生命科学名著

Textbook of Structural Biology

结构生物学

从原子到生命

[瑞典] Anders Liljas Lars Liljas Jure Piskur
Göran Lindblom Poul Nissen
Morten Kjeldgarrd 编著
苏晓东 等 译

 科学出版社

这是一本近年来出版的非常优秀的结构生物学教科书，由六位结构生物学领域国际知名科学家合作撰写。它以生物学功能为主线，以生物大分子及其复合物的三维原子结构为中心，深刻解析重要生命活动的结构基础和分子机理，内容涵盖从蛋白质、核酸、脂类到生物膜，从遗传到蛋白质产生与消亡，从信号转导、细胞运动与相互作用和演化到免疫系统和病毒结构与功能，系统深入，内容丰富翔实，图文并茂，整合了迄今最新的研究成果和学科知识，展现了从生物大分子的原子结构到重要生命活动的内在联系和基本原理。结构生物学作为一门交叉学科自20世纪90年代诞生以来，研究内容和影响日益扩大，学科内涵日臻清晰，该书及结构生物学领域近年来其他专著和教科书的不断出版，标志着结构生物学已经成为一门独立成熟的学科，正推动人们对生命现象的认识深入到以生物大分子精确三维结构为基础的新层次。

——王大成，中国科学院院士，研究员，中国科学院生物物理研究所

该书详细系统地介绍了现代生物学的一个重要分支——结构生物学，书中许多内容描述得丰富多彩、深入浅出，例如对于G蛋白结构机理及相关领域介绍的章节在其他专著中是较难看到的。总体来说，这是一本不可多得的生物大分子结构与功能百科全书式的综述专著。

——丁建平，研究员，中国科学院上海生命科学研究院生物化学与细胞生物学研究所

该书图文并茂、全面系统，是结构生物学领域近年出版的优秀专著及教科书之一，它对于结构生物学的入门者及资深科学家都具有很好的学习及参考价值。

——许瑞明，研究员，首批“千人计划”学者，中国科学院生物物理研究所

非常高兴看到《结构生物学——从原子到生命》的出版，这对于促进结构生物学在我国的快速发展非常及时。该书的作者大都是结构生物学领域的具有国际水平的专家，他们能够抽出时间和精力编写这样一本雅俗共赏、通俗易懂的专著式教科书，实属难能可贵。

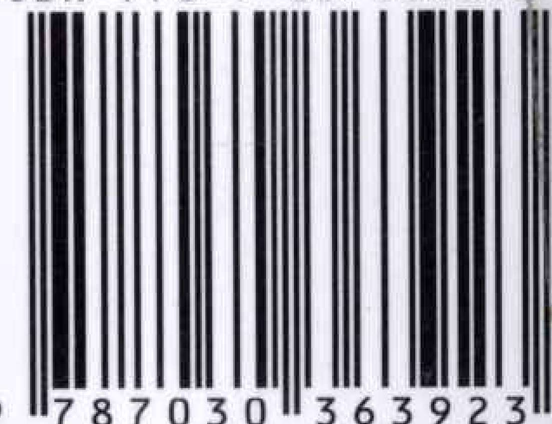
——施一公，教授，首批“千人计划”学者，清华大学医学院和生命科学学院



科学出版中心　生物分社
联系电话：010-64012501
E-mail：lifescience@mail.sciencep.com
网　址：http：// www.lifescience.com.cn

www.sciencep.com

ISBN 978-7-03-036392-3



9 787030 363923 >

销售分类建议：生物学／结构生物学

定价：168.00元

生命科学名著

结构生物学

从原子到生命

Textbook of Structural Biology

〔瑞典〕 Anders Liljas
Lars Liljas
Jure Piskur
Göran Lindblom
Poul Nissen
Morten Kjeldgarrd 编著

苏晓东 等 译



科学出版社
北京

图字：01-2011-4496 号

内 容 简 介

本书以生物学功能为主线，以生物大分子及其复合物的三维原子结构为中心，全面深刻地解析了重要生命活动过程的结构基础及由此阐发的分子机理，内容涵盖了从蛋白质、核酸、脂类到生物膜的基本结构信息及知识，从遗传信息 DNA 到 RNA 到蛋白质的传递，到基因产物蛋白质的产生与消亡，从各类生物信号转导、细胞运动、物质输运与相互作用，到免疫系统的结构生物学和病毒结构与功能，系统深入，内容丰富翔实，图文并茂，整合了迄今最新的研究成果和学科知识，展现了从生物大分子的原子结构到重要生命活动的内在联系和基本原理。

本书适合对结构生物学感兴趣的各学科、各层次的科研工作者参考， 特别适合从事与结构生物学相关领域研究的高年级本科生和研究生使用。

图书在版编目（CIP）数据

结构生物学：从原子到生命 /(瑞典) 利尔加斯 (Liljas, A.) 等编著；苏晓东等译 . —北京：科学出版社，2013.1

(生命科学名著)

书名原文：Textbook of Structural Biology

ISBN 978-7-03-036392-3

Ⅰ. ①结…　Ⅱ. ①利…　②苏…　Ⅲ. ①生物结构—分子生物学　Ⅳ. ① Q617

中国版本图书馆 CIP 数据核字（2012）第 319003 号

责任编辑：罗　静　夏　梁　刘　晶 / 责任校对：朱光兰

责任印制：钱玉芬 / 封面设计：美光制版

科学出版社 出版

北京东黄城根北街 16 号

邮政编码 :100717

http://www.sciencep.com

北京天时彩色印刷有限公司　印刷

科学出版社发行　各地新华书店经销

*

2013 年 1 月第　一　版　开本：787 × 1092　1/16

2013 年 1 月第一次印刷　印张：29

字数：685 000

定价：168.00 元

（如有印装质量问题，我社负责调换）

中译本序

古时候，西方国家的通用学术语言是拉丁语，现在主要是英语。长期以来，东亚文明（如中国和日本）的学术界主要使用自己的语言。在当前这个日趋全球化的世界，大家使用一种共同语言进行交流是很有益的，然而，要让全世界的人们使用一种通用语言变为现实还需要长时间的不断努力。因此，我们（《结构生物学——从原子到生命》的作者们）非常欣喜地看到这本教科书幸运地被同时翻译为汉语及日语。

结构生物学这门学科致力于研究生命系统中的大分子及其复合物的三维原子结构细节，已经给生物学的每一个被它提供了结构信息的分支带来了巨大革命，给相应的生物学功能提供了原子水平的分子机理和更加深入可靠的认识。这场革命始于 1953 年，DNA 双螺旋原子模型是其标志性结构，其深远影响直到今天仍在继续。通观结构生物学的发展历程，也是众多科学领域精诚合作的成功典范，物理学、化学、生物学、医学、计算机科学及其他学科的相互交叉渗透与合作成就了今天的结构生物学，使其不仅停留在学术活动层面，而且已经成为药物研究及生物技术领域的重要基础。

中国在结构生物学领域的努力可以回溯到 20 世纪 70 年代对胰岛素的晶体学研究所做的出色工作。如今，很多具有很强活力的结构生物学中心正在很多大学及研究所中以指数速度增长，遍及中华。 以中华民族的人口与技能计，我们可以确信地预见中国将很快成为结构生物学领域的领头国家。如果此书能够对此过程尽绵薄之力，我们将不胜荣幸。

Anders Liljas

2012 年 8 月 18 日于瑞典

关于作者（本书主要由 Anders Liljas 教授发起、组织及校对完成全书）：

Anders Liljas 教授是瑞典著名结构生物学家，在瑞典著名的 Uppsala 大学和 Lund 大学教学科研四十多年，曾是 Lund 大学同步辐射装置（MAX-Ⅱ）蛋白质晶体学线站的发起者及负责人。 他多年来专注于核糖体蛋白的结构与功能研究，是核糖体结构与功能及蛋白质体内翻译领域的学术权威，他主编过多本关于核糖体的专著，发表各类学术论文 100 多篇，在 *Nature*，*Science*，*EMBO J.*，*Structure*，*J. Mol. Biol.* 等国际著名科学期刊上发表过很多重要的蛋白质结构与功能研究文章，特别是与核糖体及体内蛋白质合成相关的文章。 由于他的出色工作及学术影响力，Anders Liljas 教授获得了很多荣誉及各类头衔，包括瑞典皇家科学院院士、欧洲分子生物学组织（EMBO） 委员及很多国际学术期刊编委等职。他担任过的职务中最引人注目的是诺贝尔化学奖评选委员会委员。

译者序及致谢

美国著名理论物理学家费曼（Richard Feynman）曾经说过：关于物质最重要的性质是物质都是由原子构成的（The most important thing about matter is that it’s built of atoms.）。而当代伟大的生物学家布伦纳（Sydney Brenner）则说：我认为关于生命系统最重要的事情是它们都具有基因。（I think the most important thing we know about living systems is that they’ve got genes.）我们现在知道，基因也是物质，基因及其产物的最为重要的性质仍然是它们都是由原子构成的有序三维结构，其功能是由这些在生命演化过程中形成的有序结构执行的。我们翻译的这本书就是集中讲述目前已知的基因及其产物的三维原子结构，以及利用这些原子结构我们可以怎样深入理解生命的机理与功能。

当前已确切知道的关于生命现象的起源、产生和发展过程是在地球的常温常压的环境中进行的，我们因此有理由相信原子与电子层面的精细三维结构及其动态特征是我们理解及研究生命过程与机理的最后分析极限，对于这个极限层面的全面深入理解与知识将是生物学走向系统合成研究的物质基础。地球上基因及其产物的数量众多、千变万化并且异常复杂，然而它们是可数的、有限的，随着“后基因组时代”生物科学向定量化及自动化、高通量、高效率、低成本化的不断发展，我们相信，在不远的未来，结构生物学和结构基因组学能够基本确定绝大部分基因及其产物的三维原子结构，这必将把生物学的研究和应用带向一个更新的层次及时代。我们希望通过翻译介绍这本结构生物学专著到中国，为发展我国的结构生物学研究与教学出力，为生命科学的令人振奋的发展前景做出应有的贡献。

本书的翻译及校对工作是北京大学生命科学学院结构生物学实验室的许多同学和老师两年多来共同努力的结果，由于学识及时间所限，错误与不当之处还希望读者反馈指正，我们将在以后合适的时机进行修改更正。 本书的主要翻译人员

如下：前言及第一、二章：苏晓东；第三、七章：金坚石；第四章：王开拓博士；第五、十二章：曹騣；第六、十三章：王晓君博士；第八、十五章：邢栋博士；第九、十六章：王子曦博士；第十章：刘翔宇博士；第十一、十四章：范雪新博士。附录A至C由高嵘博士翻译，附录D、E由王晓君博士翻译，附录F由苏晓东译校。校对工作主要由下列人员完成：高嵘、苏晓东、王家槐、金坚石、邢栋、曹騣、范雪新、王子曦、刘翔宇等。

我们还特别要衷心感谢北京大学生命科学学院结构生物学实验室李兰芬老师两年多来对于此书翻译工作的大力支持及组织，衷心感谢中国科学院生物物理研究所王大成院士和许瑞明研究员、中国科学院上海生命科学研究院生物化学与细胞生物学研究所丁建平研究员以及清华大学医学院和生命科学学院施一公教授对于本书的大力支持、鼓励与推荐。

译者

2012年11月2日

前　　言

结构生物学领域的第一本教科书可能要追溯到 1969 年由 Dickerson 和 Geis 撰写的《蛋白质的结构与行为》（*The Structure and Action of Proteins*），那是一本优秀的教科书，涵盖了当时已知的蛋白质结构的方方面面。当时尽管第一个蛋白质结构——肌红蛋白的晶体结构已经被解析了近十年，但全世界仍然仅有为数不多的其他几个蛋白质晶体结构被解析出来。DNA 的结构在 1953 年由纤维衍射（fiber diffraction）的数据被推演解析出来，可是关于核酸结构的知识当时还很有限。有关结构生物学初始阶段令人兴奋的研究在 RE Dickerson 最近 (2005 年) 出版的 *Present at the Flood*: *How Structural Molecular Biology came about* 一书中得以很好地讲述。现在，经历了大约 40 年，结构生物学领域有很多教科书涌现出来。例如，由 Branden 和 Tooze 编写的《蛋白质结构简介》（1991 年和 1999 年两版）是一本非常出色的教科书，但其缺乏最近 10 年间（以此书的出书年份推算——译者）令人激动的结构生物学最新进展。目前，在蛋白质数据库（PDB）中的结构信息已经非常庞大并且仍在迅速广泛地增长着。这种增长速度很大程度上源于技术上的飞速发展，其中的关键是运用同步辐射作为 X 射线光源。结构生物学领域的发展已不再单纯是学术研究的努力结果，长期以来与制药及生物技术相关的工业界也对其做出了巨大的贡献。同时，结构基因组学实验室的发展使得实验方案进一步完善，每一个实验者都可以从中获益，从而使得蛋白质结构的测定稳步地提升。

面对这些如洪水般涌来的信息，必须做出选择。我们的选择是基于我们的教学经验，也就是那些学生们最能从中获益的信息。关于蛋白质、核酸和膜结构的核心知识是通过与生物学的中心法则联系在一起被描述的。我们的研究领域涉及结构生物学中较广的范围，我们自己的研究侧重点也会对这些选择有一定的影响。

由于最近的研究表明膜脂在细胞中行使重要并且主动的生理功能，例如对膜上的酶和通道的调控以及脂筏的形成发挥着重要的作用，所以本书专门用一章较为全面地讨论脂类的物理特性及功能。

对于出版这样一本结构生物学教科书，我们得到了许多同行们慷慨的帮助，诸如审阅文稿、提出专业性观点、同意我们引用他们已发表或一些尚未发表的成果等。我们无法将他们一一列出，生怕会有所遗漏。我们对所有这些帮助表示衷心的感谢。我们特别要对 Terese Bergfors 表示感谢，感谢她非常认真地阅读文稿并对内容及语言提出的修改。书中大多数的插图是用图形程序 Molscript、Molray 和 PyMOL 来绘制的。Molscript 和 PyMOL (http://pymol.sourceforge.net/) 是较早期的以图示的方法描述蛋白结构的程序 {http://www.avatar.se/molscript/molscript.html, Kraulis, PJ (1991). Molscript: a program to produce both detailed and schematic plots of protein structures J. Appl. Cryst. 24, 946-950，Molray [Harris M, Jones TA (2001). Molray: a web interface between O and the POV-Ray tracer. Acta. Cryst. D57:1201-1203]}。 我们对获得许可使用这些程序表示衷心的感谢。

目　录

第一章

导　　论

1.1 生命

我们的周围充满了微生物、植物与动物这些我们马上就能够识别的生命体（图 1.1），但是给生命下一个确切定义并非易事。就本书的目的来说，或许可以将生命定义为一个具有化学活性且能够进行繁殖及演化的单元。

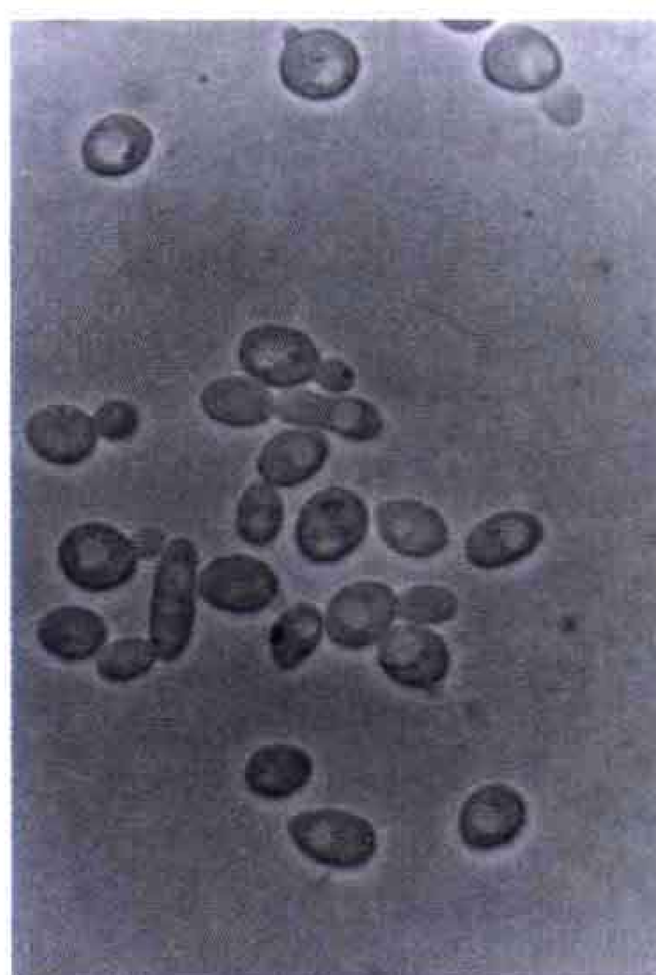






图 1.1　生命体具有很多种不同形式。上左：面包酵母细胞的显微图片（由 Concetta Compagno 提供）；上右：林奈花（*Linnea borealis*）覆盖着拉普兰的广阔区域（由 Bernardo Rotar 提供）；下：麋鹿，斯堪的纳维亚最大的陆地动物（由 Aca.Pixus.dk 提供）。在这些不同外部形式的内部存在着非常类似的分子结构，决定着生命载体的生活方式。

化学活性即代谢，涉及能量与物质的转换。代谢活性使生命机体以不同的形式获得能量与化学物质。上千种化学活性可以在活着的机体中同时发生，这些反应必须很好地协调以保证生命单元的稳定性。

这些生命单元在繁殖（产生新的单元）过程中同时产生延续性及变异性这样的重要生命特征。繁殖、信息的水平转移及“出错”复制的组合构成了演化的基础。换句话说，生命单元的组成应该能够随时间变化以更好地适应变化中的环境条件。生命机体看起来有着非常不同的外形及生活方式，然而，生命的基本特征包括的代谢、繁殖和演化却是由非常相似的亚结构、生物大分子及细胞组成和掌控的。

1.2 生命的组织层次

生命世界由小到大具有几个等级层次。最底层是分子，是无机化合物、有机化合物及生物大分子的混合体，然后是亚细胞结构、细胞、组织、器官、机体、种群、群体及生物圈，这些构成了地球上的所有生物群。

大分子是所有生命机体的核心分子。它们是由相同或不同单体的单元成分构成的巨大多聚物（图 1.2），以共价键或者非共价键相连。大分子执行很多功能，如能量或者信息的储存、反应的催化、协作及调控、通讯、结构支持、防御、运动及运输，这些都是代谢、繁殖和演化的基础。大分子按照其化学组成不同可以分为 4 种类型：①多肽及蛋白质是由氨基酸残基组成的多聚体；②核酸是由核苷酸组成的多聚体；③碳水化合物是由糖组成的多聚体；④生物膜是由脂类聚集成的结构。在以下的章节中我们将讨论生物大分子的结构，并且将它们的功能与更高层次的生命世界相联系。

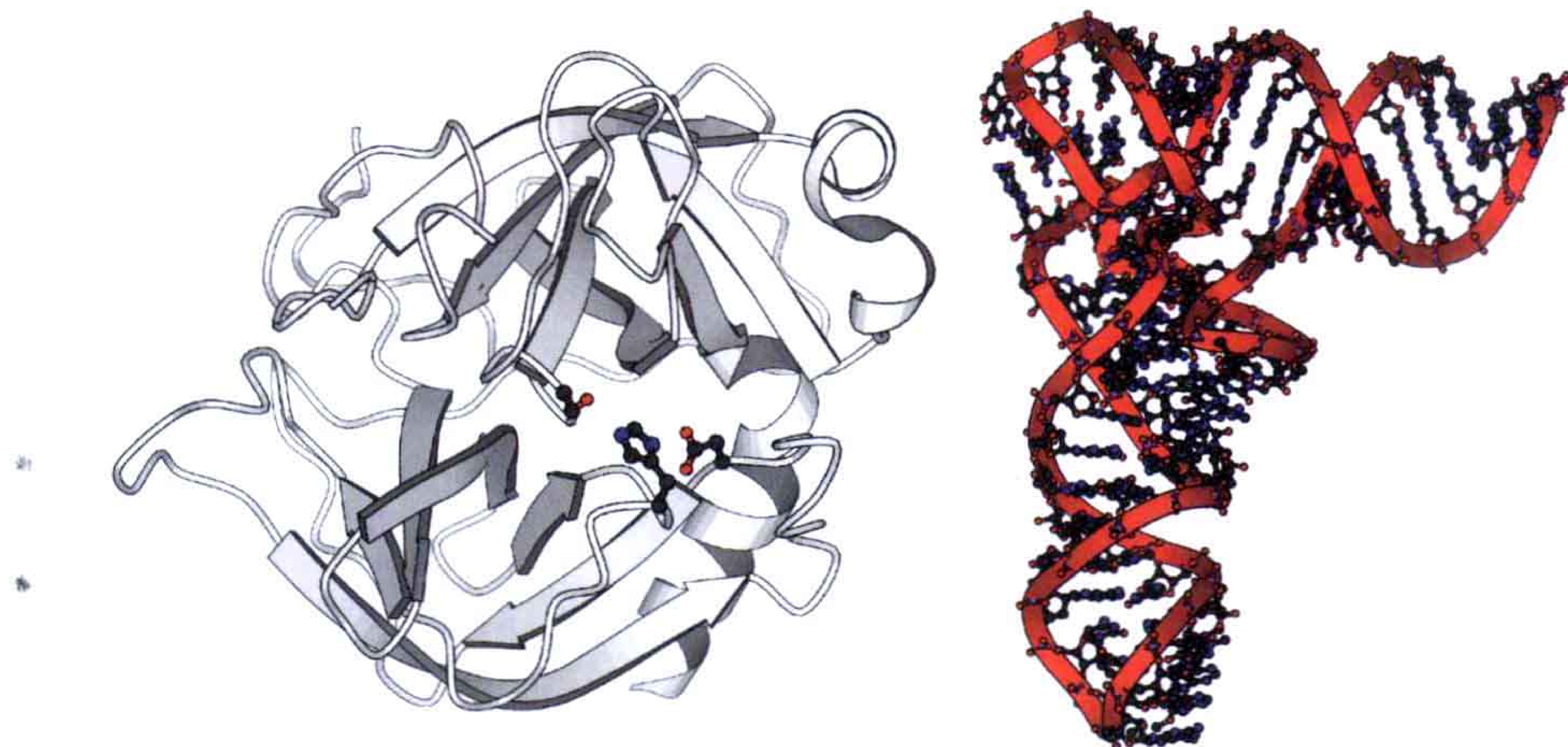


图 1.2　两个大分子的简化结构图。左：一个非常著名的蛋白质结构（糜蛋白酶，PDB：4CHA）；右：一个核酸分子的结构（酵母的 $tRNA^{Phe}$，PDB：1EHZ）。它们在以后的章节中还会专门提到。

生命的基本单元是细胞。细胞被质膜包围，从而与外界环境分离并且产生分开的空间以保持相对恒定的内部环境。细胞一般具有两种组织特点：①细菌和古细菌的原核特征；②真核生物的真核特征。原核细胞一般比真核细胞要小，平均大小为1μm×1μm，它们一般以单细胞形式存在。所有原核细胞都具有同样的基本结构，即质膜、细胞内包含DNA的类核体及被称为胞质的其余胞内材料，其中含有核糖体、许多酶及细胞骨架元素。真核细胞一般比原核细胞至少大10倍，也更加复杂。在真核细胞中，质膜及胞质在内的基本的原核细胞结构被升级为各种间隔，这些间隔也被称为细胞器。在真核细胞的胞质中，可以发现更多的特殊结构。在一些情形下，这些细胞器内部被膜彼此分开。最为常见的细胞器有：①细胞核，存储细胞的遗传物质并且是复制及基因表达的场所；②胞质，蛋白质合成及许多重要生化反应的发生地；③线粒体，“发电厂”及能量储存区间；④内质网及高尔基体，蛋白质打包及进一步分送的场所；⑤溶酶体及液泡，蛋白质等多聚大分子降解为单体的场所（图1.3）。

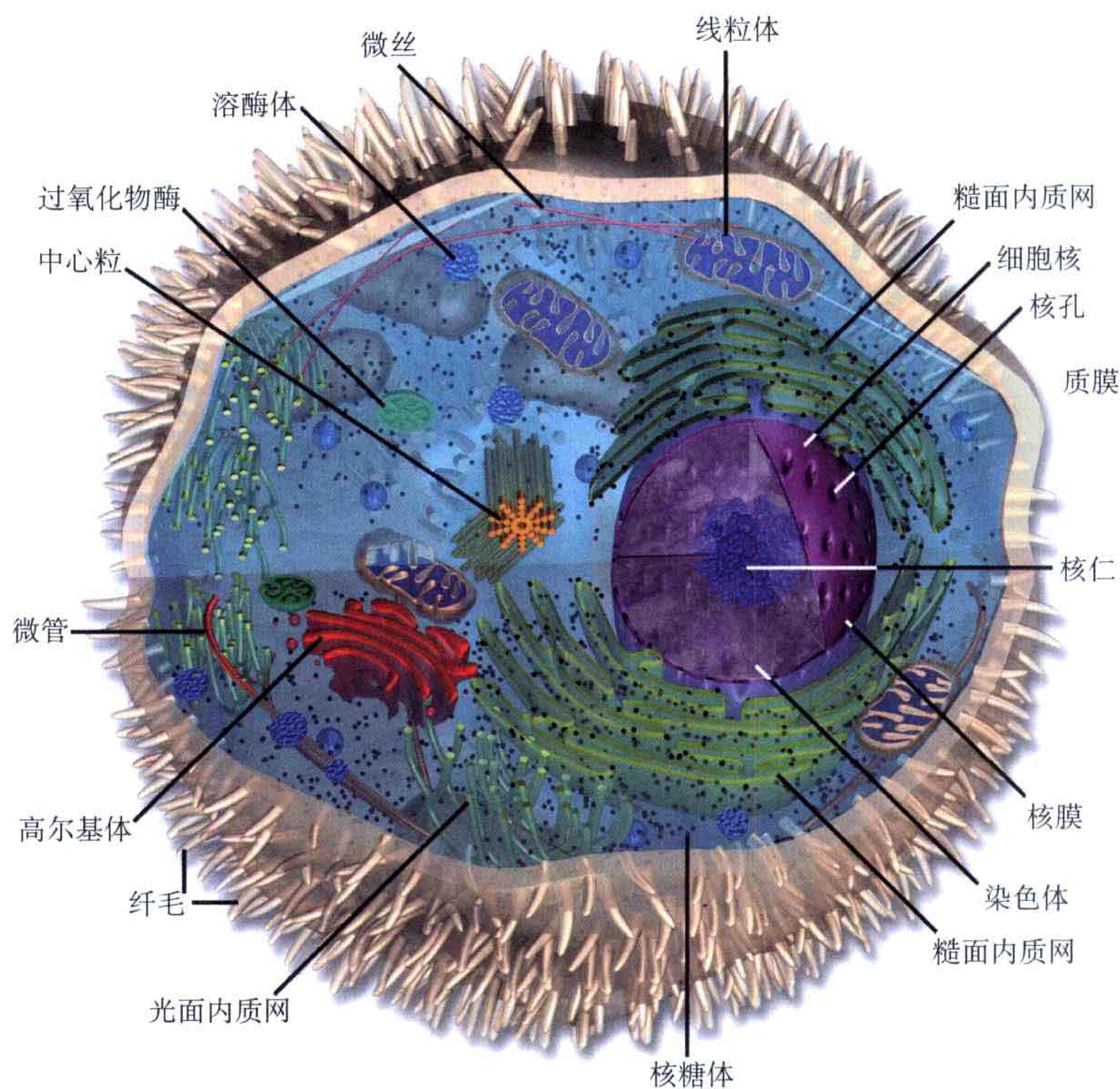


图1.3　动物细胞简图。展示了如细胞核、膜系统（ER）及线粒体等亚细胞结构。

人们相信地球上的所有生物体都起源于一种（原始的）单细胞生物。今天，我们已发现几百万种不能够互相交配繁殖的生命机体，它们被称为物种。这些物种都是完美的，因为它们都成功适应了其所处的各种不同环境。然而，明天有些物种可能会不那么完美并且因适应不了环境而灭绝，有许许多多以前广泛分布在地球上的物种已经

灭绝。简而言之，生物体（物种）的多样性是一个非常动态的过程，随着时间地点的不同而不断变化着。

导致现在生物多样性的事件及曲折过程可以被形象地表述为一个“演化树”，它展示了一个物种分叉演化成为新物种的（时间）顺序。这棵“树”记录了生活在过去不同时间的先祖演化为后代的足迹。换句话说，“演化树”显示了现代及古代物种间的演化关系。当我们比较不同物种中相似生物过程的大分子结构时，理解这些物种间的演化关系是至关重要的。由于没有足够的化石记录，一些早期的分支是很难重构出来的。然而，基于现代生命机体的分子证据，我们可以把所有活着的机体分为三个“大界”（domain），它们在十多亿年以前就已经分离演化出来：①古细菌；②细菌；③真核生物（图 1.4）。尽管表面看来古细菌和细菌很相似，其实它们在演化历史的很早时期就已经分成迥然不同的家系了。

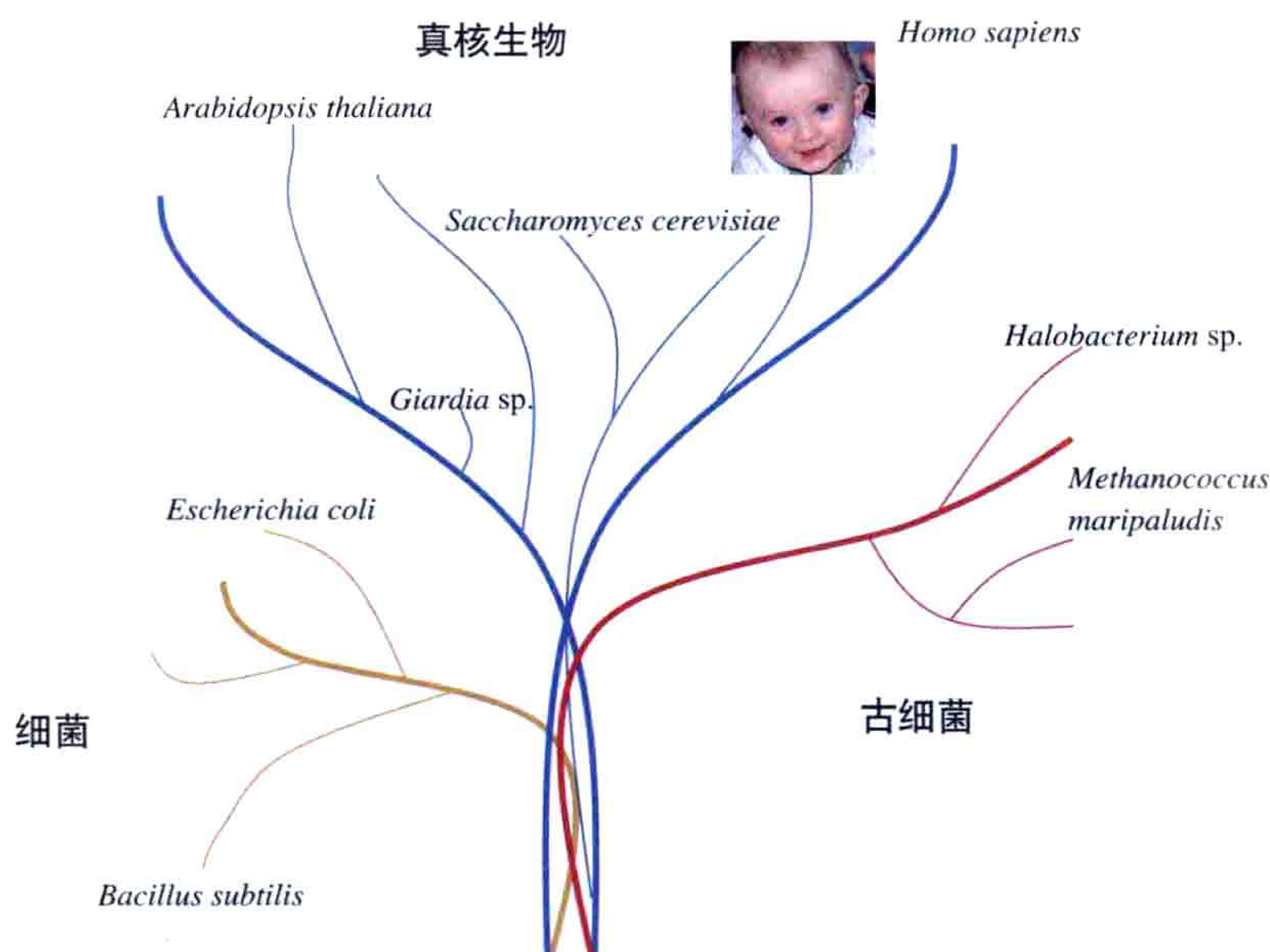


图 1.4 简化的生命树。生命的共同祖先起源于大概 40 亿年以前。第一个分叉的具体位置介于细菌、古细菌和真核细胞之间，仍然不清楚。

古细菌是首先在极端环境如酸性或碱性温泉、海底、盐井中发现的生命体，但是它们也可以在比较“正常”的环境中被找到。它们的复制、转录及翻译机器与真核生物的相应机器相似，但是古细菌的代谢及能量转换系统与细菌的相应系统更类似。

细菌由十多种子群组成，也被称为演化枝（clade），最重要的子群有：原始菌族（Protobacteria）、蓝细菌族（Cyanobacteria）、螺旋菌族（Spirochete）、衣原体菌族（Chlamydia）及硬壁菌族（Firmicute）。原始菌族是最大且非常多样化的子群，包括研究得最为深入的物种——大肠杆菌（*E. coli*）。有时，细菌按其细胞壁的组成分为革兰氏阳性菌（Gram-positive bacteria）如枯草芽孢杆菌（*Bacillus subtilis*），以及革兰氏阴性菌(Gram-negative bacteria)如大肠杆菌。细菌在生化代谢上展现出最大的多样性。

真核生物可以被分为四大类：原生生物界（Protista）、植物界（Plantae）、真菌界（Fungi）及动物界（Animalia）。原生生物界主要包括单细胞生物，且具有多重起源，其中有一些是非常初级的真核生物，如鞭毛虫（*Giardia*）；另一些则与动物类如盘基网柄菌属（*Dictyostelium*），或者植物类如红藻（red algae）联系密切。吞噬（phagotrophy）是一种由质膜形成小袋来包裹“食物”的取食方式，是真核生物的主要特征。

1.3 地球上的生命简史

化学演化理论认为，至少在40亿年以前原始地球的条件使得第一批生物分子出现。奥巴林（Oparin）和豪尔丹（Haldane）在20世纪20年代各自独立地提出，如果地球的最早期大气是还原性的并且有外界能量输入，则一系列的有机化合物是可以被合成出来的。20世纪50年代斯坦利·米勒（Stanley Miller）和哈罗德·尤瑞（Harold Urey）在实验室里模拟了这些条件，将水蒸气、氢气、氨气和甲烷气体混合在一起并且给予其电击，几天以后在此系统中可以观察到好几种复杂分子如氨基酸、核酸等这些构成现代生命的基本物质出现了。得到这些单体以后即使在非生命条件下进行多聚化也并不困难。然而，最早的多肽及核酸是怎样“活起来”的？换句话说，它们是怎样开始繁殖和演化的呢?

术语“复制子”是指一种结构，这种结构仅当周围预先存在同种结构时才能形成。例如，一个微小晶种引入过饱和的盐溶液时会引起结晶。然而，这个例子仅表示基于简单结构的一个简单“复制子”，更复杂的“复制子”应该存在几种可遗传的简单变异形式。在可持续的演化过程中，遗传需要无限多的结构形式和无限多的变异。第一个人造“复制子”是一个无需酶来复制（由单体聚合而成）的脱氧六核苷酸，由基德罗斯基（von Kiedrowski）在1986年合成。现在，我们相信最早的短链RNA分子能够催化其子代分子的多聚。最早的活RNA分子（即RNA世界——译者注）需要同它自己的其他错误拷贝及低效率的复制系统竞争，以成功地得到复制所需的单体。即使能够自我复制的RNA分子满足了上述生命起源的需要，通往第一个细胞的道路仍然是艰难曲折的。其中的主要步骤是引入蛋白质从而建立核酸 - 蛋白质世界，再引入膜系统从而与外界环境分离形成原始细胞。

作为目前所有生命体共同祖先的第一个细胞，其起源可以追溯到大约40亿年以前，那时，复制与翻译复杂机器已经存在。有一个假说在随后的20亿年中被提出，单细胞系统已经演化到具有精细的代谢反应网络，联结越来越多的复杂分子机器进行核酸复制和RNA到蛋白质的翻译，还具有维持细胞较大的可塑性以适应复杂多变的外界环境要求。在这段时期，初始活细胞的主要能量来源仍然依赖于非生物起源的有机化合物。后来，大约25亿年前，另一个演化过程的重大步骤是出现了能够利用太阳光作为代谢能量的细胞。光合作用提供了独立的能量来源，并且很快产生了大量的有机物质和氧气。

有氧代谢的引入显著地改变了细胞的生化反应。许多酶反应直接或间接地变得依赖于氧。有氧代谢使得细胞长得更大。其他一些重大演化步骤包括两性的起源、多细胞生命的起源和社会群体的起源。所有这一切演化事件的背后都是蛋白质、核酸、糖类及脂类的结构与功能的演化过程。

1.4 结构生物学是什么，始于何时？

结构生物学领域集中在一个经典共识：要真正理解某事，我们需要看见它（的结构），也可以说为“眼见为实”。这个共识对于宇观研究对象如天文学与天体物理、宏观研究对象如设计机器及修理汽车与钟表、微观研究对象如生物系统或粒子物理都是如此。结构生物学是试图使生物学中的亚细胞及分子客体（结构）可视化并以此来理解机理的科学。

在1964年的一篇综述中，结构生物学的先锋之一詹姆士·华生（James Watson）指出：“不幸的是，除非我们首先知道其结构，否则不可能在化学水平精确描述一个分子的功能”。这的确在很大程度上点中问题的要害，这个观点不断被生物学发展所证实，证实这个远见卓识的一个经典案例是华生在说此话的十多年前解决了DNA的双螺旋结构。在弗朗西斯·克里克（Francis Crick）和詹姆士·华生利用罗萨林德·富兰克林（Rosalind Franklin）与毛瑞斯·维尔金斯（Maurice Wilkins）测得的衍射数据搭建出DNA结构模型之前，核酸在分子生物学中的中心地位是没有被正确理解的。（双螺旋结构）模型立即提供了解释DNA自我复制、DNA转录到RNA及蛋白质翻译的主要步骤的机理。

要精确指出结构生物学的起点是困难的，其发展的一个重要步骤是提纯大分子的组分。弗里德里希·麦舍耳（Friedrich Miescher）于1869年就发现并且分离了DNA，而对其生物学功能的理解则一直要到1944年在埃沃瑞（Avery）、麦克列奥（MacLeod）及麦克卡提（McCarty）提出DNA是遗传物质以后才获得。1953年对DNA结构的揭示是结构生物学的一个重大里程碑。

人们很早就已经知道蛋白质了，但是一直不太清楚其分子本质。蛋白质最初被当成“胶体”，被认为没有确定的结构及形状。酶被当成是与脂类、糖类及蛋白质很不同的物质，只在动植物中以很低的浓度存在。提奥多·斯韦博格 [Theodor (The) Svedberg] 利用他发明的超速离心法展示了蛋白质具有确定的分子质量，这一发现成为理解蛋白质本性的重要步骤。J.B. 萨姆纳（Sumner）和后来的J.H. 瑙斯儒珀（Northrop）提纯并结晶了脲酶及胃蛋白酶，从而显示了蛋白质具有确切独特的结构。与此同时，W.M. 斯坦利（Stanley）结晶了好几种病毒。完善结晶及利用晶体学方法分析晶体结构经历了几十年的发展，直到1959年和1968年分别解析出了肌红蛋白和血红蛋白的晶体结构以后才成熟起来。

除衍射和散射以外，结构生物学还包括其他几种研究方法。电子显微学在早期就

已经成为提供生物系统和大分子组织情况的重要技术。这方面的一个重要进展是卡斯帕（Caspar）和克鲁格（Klug）在20世纪50年代末、60年代初分析病毒颗粒的结构。在其研究中对称性原理能够被推导出来，从而确定较大病毒中不同的功能组分。在电子显微学领域的另一个重大进展是R. 翰德森（Henderson）及N. 昂文（Unwin）利用电子衍射研究了细菌视紫红质（bacteriorhodopsin）的二维晶体结构。尽管这些工作开辟了结构研究的崭新的可能途径，但其只能对于有限的能得到足够好材料的研究对象进行结构研究。随后经过进一步发展，大分子复合物的单颗粒重构方法和断层重构法（tomography）大大扩展了电子显微学在低分辨率及中等分辨率下对结构生物学的贡献。特别是单颗粒重构方法能够确定样品中的几种不同构象，从而得到功能分子的动态图像。

结构生物学对很多生物系统的理解已经从表征大分子团块推进到原子坐标及详细的分子相互作用层次。

从原则上讲，晶体学的一个局限是给出的结果是研究系统的静态图像。在幸运的情况下，好几种不同状态可以被结晶及表征到原子分辨率，这通常会将功能活性的可能性假说限制到几种情况，有时是唯一的。然而，很多情况下我们需要得到晶体学不能获得的功能状态（或许因为这些状态寿命太短），或者需要了解系统的动力学信息。在这方面，核磁共振（NMR）研究有时可以提供缺失的信息。

一般来说，核磁共振谱可以得出蛋白质结构。如果该蛋白质得不到晶体，NMR结构信息就非常宝贵；如果NMR及晶体结构都可以得到，晶体学信息的质量通常要更高。但是，核磁共振谱学的独特贡献来自于当结构已知时对系统的动力学研究，在这里，运动及瞬时相互作用细节可以被刻画出来。

要想得到最完整的信息，应该同时利用几种方法。这样，错误或片面信息可以被纠正或扩展。在理解得最好的生物学系统中，物理学家及理论化学家会从更多的角度在实验或计算方法上对理解这个系统作出贡献。

1.5 对本书的简短介绍

这是一本面向大学本科高年级学生及研究生的结构生物学教科书。本书的焦点集中在对生物学中的重点及有趣方面进行结构阐述，强调的是利用结构信息理解生物学。本书并没有试图涵盖生物学或者分子生物学的所有方面，而是将重点放在那些结构信息了解得较多的系统上。

本书覆盖了蛋白质、核酸、脂类及膜领域的基础知识，包含了很多隐含在原始文献中但对于结构生物学相关领域的工作者非常重要的观察事实。对于需要复习基础化学知识的学生，我们推荐先读一读相关附录。附录中还包含了一些利用蛋白质序列及结构相关数据库的必需工具，以及一个简短的蛋白质及结构生物学相关的诺贝尔奖得主的介绍。

本书并不试图描述得到结果的方法，这需要不同的出版物共同完成。然而，这本内容有限的教科书向我们综合地描述了蛋白质、核酸、脂类及生物膜的结构与功能。很遗憾没能包括多糖生物学，生物学中多糖材料的复杂性超过了本书描述的任何其他类分子，因而需要更多的空间。况且，目前多糖系统的结构知识还是相当有限的。

除了有关蛋白质、核酸及脂类的基本结构知识外，本书侧重于表述有关 DNA 中的遗传信息表达到蛋白质的各个步骤。同样，大分子的降解过程也得以详述。对与蛋白质聚集相关疾病的理解最近有了很大的进展，本书中各相关章节有所涉及。很多生物学功能与细胞膜有关，它们包裹着细胞及细胞器。物质与信息的跨膜传输对于所有的细胞生物学过程都至关重要，近来这方面的结构知识发展得很快。在多细胞生物中，细胞 - 细胞的接触与相互作用对于协调多细胞活性是非常基本的，与此相关联，细胞及机体是怎样运动的也被了解得越来越充分，这些领域都在本书中有详细描述。

对于生物系统演化及功能基因组学的深入理解也得益于结构生物学研究。完整基因组的 DNA 序列测定正在以前所未有的速度增长，产生海量的数据库，序列同源性（identity）比较高的 DNA 或者蛋白质序列能够很可信地被确认出来。在很多情况下，序列本身不足以说明问题，然而，结构的相关性比较能够提示一个蛋白质的演化关系及其功能。

（苏晓东 译校）

第二章
蛋白质结构基础

2.1 蛋白质的构象

大部分生物过程中，从分子、细胞或者机体水平上看，蛋白质是主角，因此，理解蛋白质功能非常重要。蛋白质的结构决定其功能。大多数蛋白质被设计为结合其他蛋白质、DNA、RNA 或者其他类型的分子，这需要蛋白质形成并且保持其功能基团精确的空间组成形式。尽管大部分蛋白质具有确定的构象，但它们不是刚性的。各种类型的构象柔韧性（侧链及环区的运动、结构域的转动等）常常对功能的影响是关键的，尽管大部分蛋白质都参与动态过程，但上述的构象柔韧性是纯粹的结构特有性质。

蛋白质的可塑性极高。它们可以具有非常不同的形状及构型，一个蛋白质序列中几乎所有位置都是可变的，同时还能够保持其结构与功能。蛋白质可以是球状或纤维状的、僵硬的或富有弹性的，它们可以进行或精妙细微、或差别巨大的构象变化。它们可以是酶，既合成或大或小的各种分子，也可以催化降解这些分子；它们也可以是分子马达，产生旋转或平移运动。在本章及后续的章节中，我们将详细描述蛋白质的这些能够使它们具有不同功能角色的结构特征。

2.1.1 氨基酸

蛋白质是由 L- 氨基酸残基（L 代表 *levo* = 左）所组成。氨基酸的另一种相反的手性被称为 D- 构型（D 代表 *dextro* = 右）。除了甘氨酸，其他氨基酸的 α 碳都是手性的，α 碳与一个氢原子、一个氨基、一个侧链及一个羧基成键。从一个 L- 构象的氨基酸氢原子看过去，顺时针可以看到羰基（CO）、侧链（R）和氨基（N）基团（即“CORN”法则，图 2.1）。

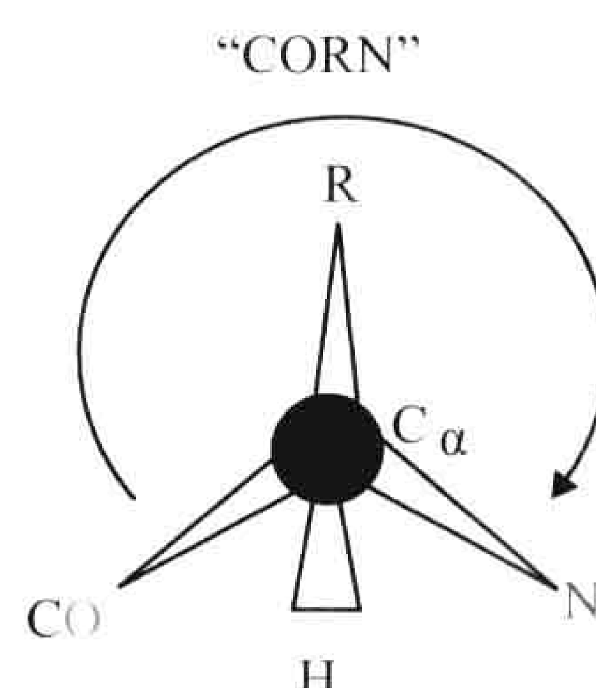


图 2.1 从 H-Cα 键向下看 α- 碳，L- 构型氨基酸“CORN”法则的示意图。

氨基酸序列是蛋白质的初级结构。演化过程中，20 种氨基酸被选择为组成蛋白质的标准集合（图 2.2）。还有两种氨基酸（硒代半胱氨酸和吡咯赖氨酸）也是遗传编码的，但是很罕见。许多氨基酸的翻译后修饰（PTM）被发现并确认（表 2.1）。这些翻译后修饰常具有重要的信号转导功能。

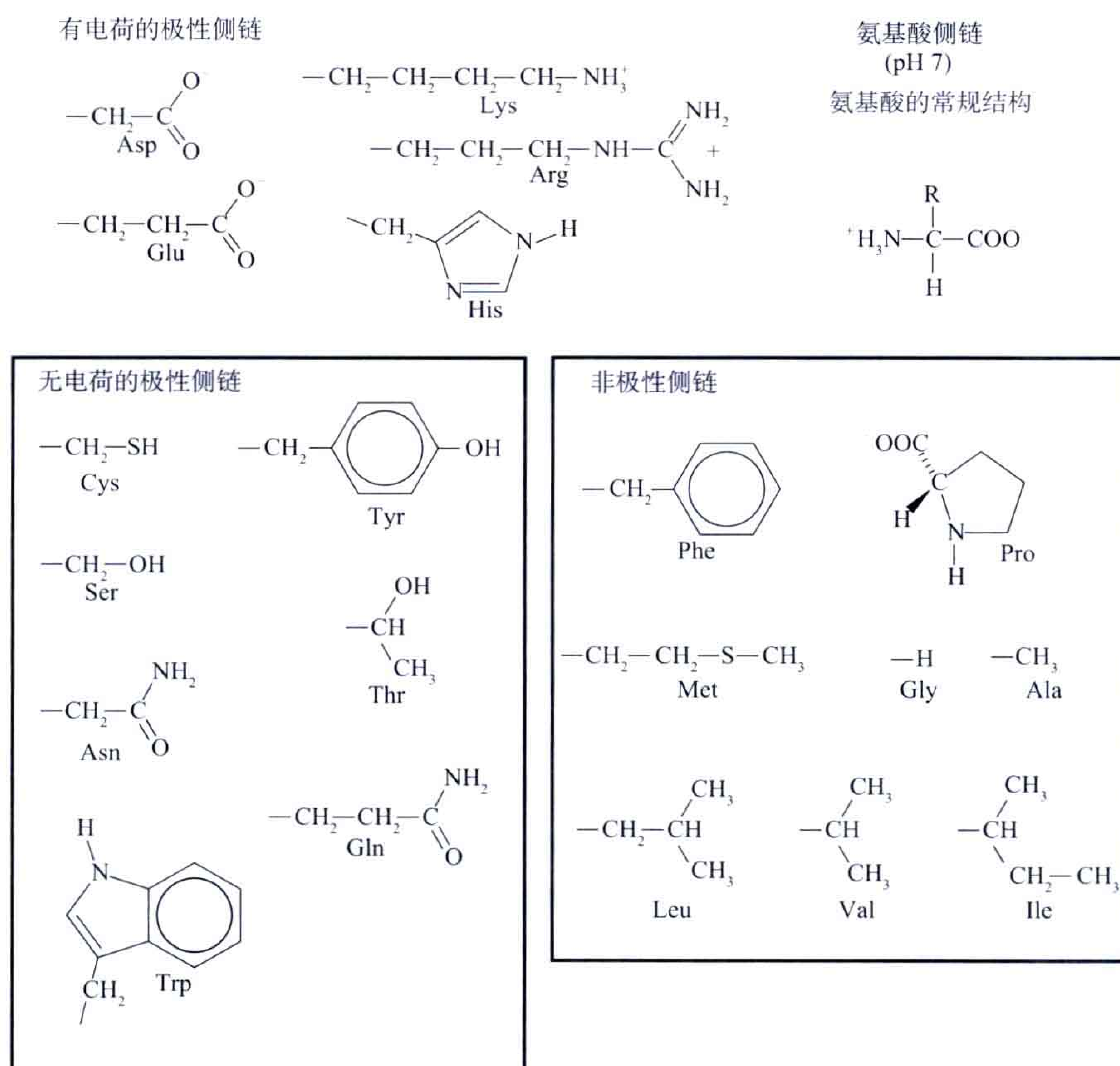


图 2.2　20 种氨基酸的不同侧链。

根据它们的侧链性质，氨基酸可以被分为几类。下面是其中一种分类的方法。

非极性侧链：	丙氨酸（Ala）、缬氨酸（Val）、亮氨酸（Leu）、异亮氨酸（Ile）、甲硫氨酸（Met）、脯氨酸（Pro）和苯丙氨酸（Phe）。
有电荷的极性侧链：	天冬氨酸（Asp）、谷氨酸（Glu）、组氨酸（His）、赖氨酸（Lys）和精氨酸（Arg）。
无电荷的极性侧链：	丝氨酸（Ser）、苏氨酸（Thr）、半胱氨酸（Cys）、天冬酰胺（Asn）、谷氨酰胺（Gln）、酪氨酸（Tyr）和色氨酸（Trp）。
无侧链：	甘氨酸（Gly）。

上述分类法并不总是有用的。例如，大的芳香环类侧链 Tyr 和 Trp，以及 Cys 基本

上是疏水的，尽管它们可以形成氢键（附录 A）；His 在 pH 低于 7 时是带正电荷的；Pro 具有很特殊的性质，它的侧链是和主链氮共价连接的。

成对的 Cys 残基能够形成二硫键，它们共价连接同一或者不同的蛋白质从而增强结构的稳定性。二硫键存在于 SH（巯基）可被氧化的环境中，如真核细胞的内质网中（ER）。

主链上 N 端的氨基和 C 端的羧基在中性 pH 时通常是带电的。它们的 pK_a 分别为 8 和 3。主链两端及氨基酸侧链所带的真实电荷对于功能是很重要的。

2.1.1.1　侧链及其相互作用

氨基酸侧链的序列给出各个蛋白质的独特性质。这些性质不仅决定了这个蛋白质的折叠，而且还决定其表面性质，从而对与其他分子的选择性相互作用及化学催化反应至关重要。

根据氨基酸在蛋白质结构中所处的具体环境，表 2.1 给出的 pK_a 值可以有很大的变化。例如，如果两个羧基基团彼此靠得很近而且没有其他正电荷基团来平衡它们的负电荷，它们的 pK_a 可以被显著地提高。同理，除通常所说的带电荷的精氨酸、赖氨酸、天冬氨酸及谷氨酸以外，半胱氨酸、组氨酸、丝氨酸、苏氨酸及酪氨酸的侧链有时在生理环境下也会带电。

侧链可以彼此间或者与主链以很多种方式相互作用（附录 A）。主要位于蛋白质内部的非极性或者疏水侧链，与其他非极性侧链相互作用。极性侧链可以以氢键的形式与其他侧链或主链相互作用。带电基团常与蛋白质表面上相反电性的侧链相互作用。被氧化的半胱氨酸可以相互反应形成共价的二硫键。

表 2.1　氨基酸基本信息

名称	三字母简称	单字母简称	侧链的 pK_a	酶促转录后修饰的例子（附录 E）
丙氨酸	Ala	A	—	—
精氨酸	Arg	R	12.5	—
天冬酰胺	Asn	N	—	糖基化
天冬氨酸	Asp	D	3.9	—
半胱氨酸	Cys	C	8.2	二硫键，异戊烯基
谷氨酸	Glu	E	4.1	—
谷氨酰胺	Gln	Q	—	—
甘氨酸	Gly	G	—	—
组氨酸	His	H	6.0，14.5	磷酸化
异亮氨酸	Ile	I	—	—
亮氨酸	Leu	L	—	—
赖氨酸	Lys	K	10.5	甲酰化、乙酰化、泛素化及 SUMO 化
甲硫氨酸	Met	M	—	—
苯丙氨酸	Phe	F	—	—
脯氨酸	Pro	P	—	—
丝氨酸	Ser	S	14.2	糖基化，磷酸化

续表

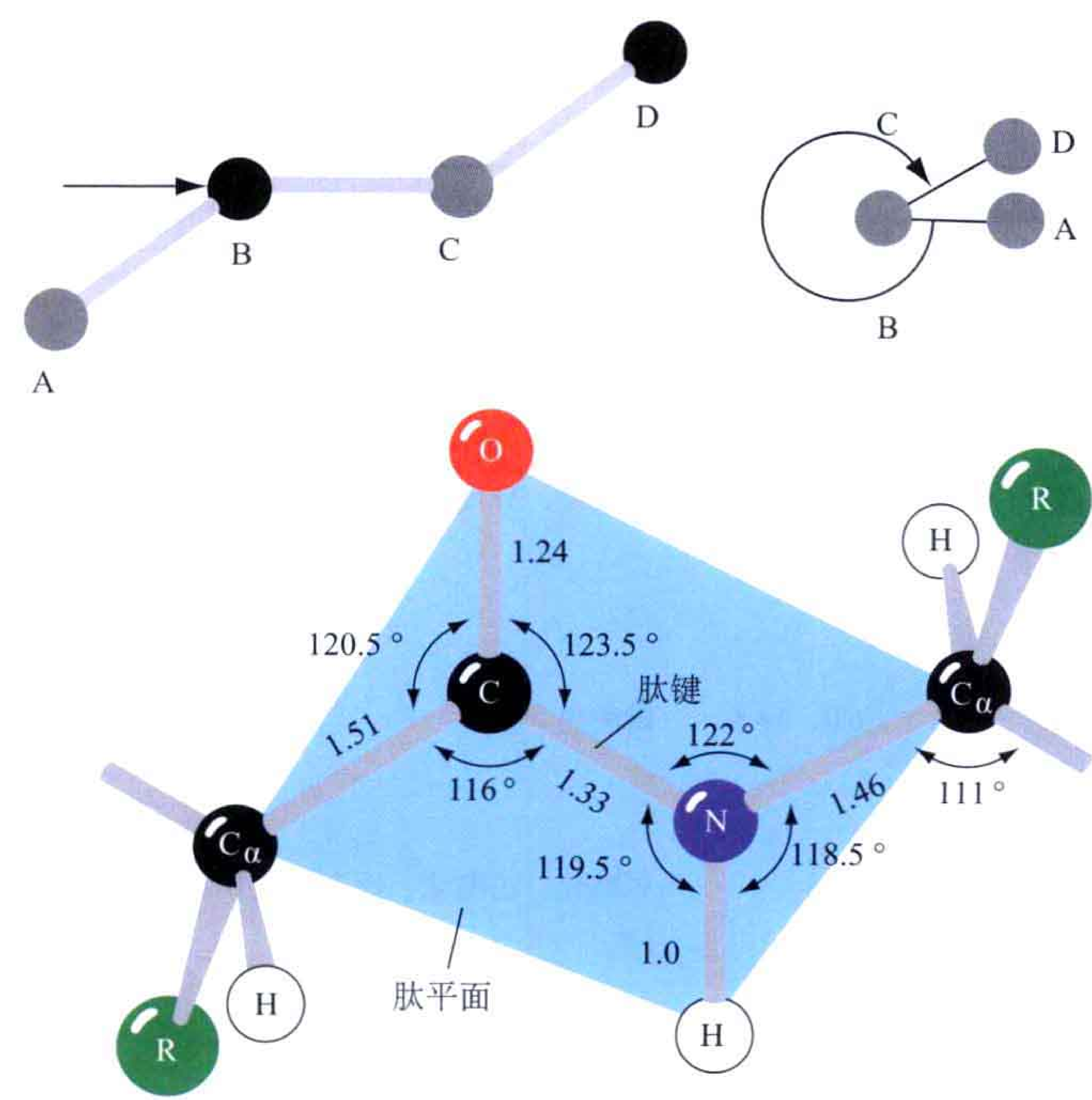
名称	三字母简称	单字母简称	侧链的 pK_a	酶促转录后修饰的例子（附录 E）
苏氨酸	Thr	T	15	糖基化，磷酸化
色氨酸	Trp	W	—	—
酪氨酸	Tyr	Y	10.5	磷酸化、硫酸化
缬氨酸	Val	V	—	—

组氨酸可有几种不同的质子化状态。咪唑环侧链上的两个氮原子都可以被质子化或去质子化。咪唑环上两个氮原子任意一个带有一个质子时是电中性的。当两个氮原子都被质子化时，咪唑环带正电荷。罕见情况下，两个氮原子都去质子化时，咪唑环会用两个氮原子桥联两个金属离子。在铜锌超氧化物歧化酶（SOD）中就是如此，组氨酸侧链的咪唑环居中，像桥一样连接两个金属原子。此时组氨酸侧链的净电荷为 − 1。

氨基酸侧链会以多种方式被修饰，特别是赖氨酸（甲酰化、乙酰化、磷酸化、泛素化及 SUMO 化）。体内氨基酸侧链的酶促或非酶促修饰在附录 E 中有进一步的讨论。

2.1.2 蛋白质的主链

在蛋白质中，氨基酸残基由肽键相连。 肽键中，由于 C 原子和 N 原子之间的单键 - 双键共振，CO 和 NH 基团之间的键具有部分双键特征。这意味着所有连接两个 Cα 原子间的肽键原子位于一个平面上。ω(omega) 扭转角 (图 2.3) 对于正常的反式（*trans*）构象肽键约为 180°。由于空间位阻作用，顺式（*cis*）肽键（$\omega \approx 0°$）非常罕见。



H R H Cα N C O Cα R H φ ψ ω

图 2.3　肽键。左上角图示扭转角（torsion angle）的定义；中上图显示当我们顺着原子 B 和 C 间的键看时怎样确定化学键 AB 和 CD 间的夹角；右上图示各扭转角的名称。在一般反式肽链中，$\omega \approx 180°$，肽键中的原子倾向于形成一个平面（酰胺平面）；下：一个肽键中不同原子间的距离（用 Å 表示）及角度。

在 Pro 残基前，顺式构象的肽键会常见一些（顺式脯氨酸），这在 9.1.1.1 节中会进一步讨论。

由于肽键的平面性，多肽主链每个残基的构象都可以用两个扭转角来描述。这两个角度是 φ 和 ψ，由于羰基氧、NH 基团的氢和 C 上的氢及侧链原子之间的空间位阻，φ 和 ψ 只能取有限的值。这点被用来定义 φ 和 ψ 的允许区域，并用拉氏图（Ramachandran plot）（图 2.4）来表述。拉氏图中两块主要的允许区域对应于蛋白质结构中观察到的两种主要构象类型（右手 α 螺旋及 β 片层）。一个较小区域对应于左手螺旋构象，其也是被允许的。对于没有侧链限制的甘氨酸（Gly）残基 φ 和 ψ 的允许区域要大得多。另外，脯氨酸（Pro）具有非常受限的构象，其 φ 的取值非常接近 −60°。

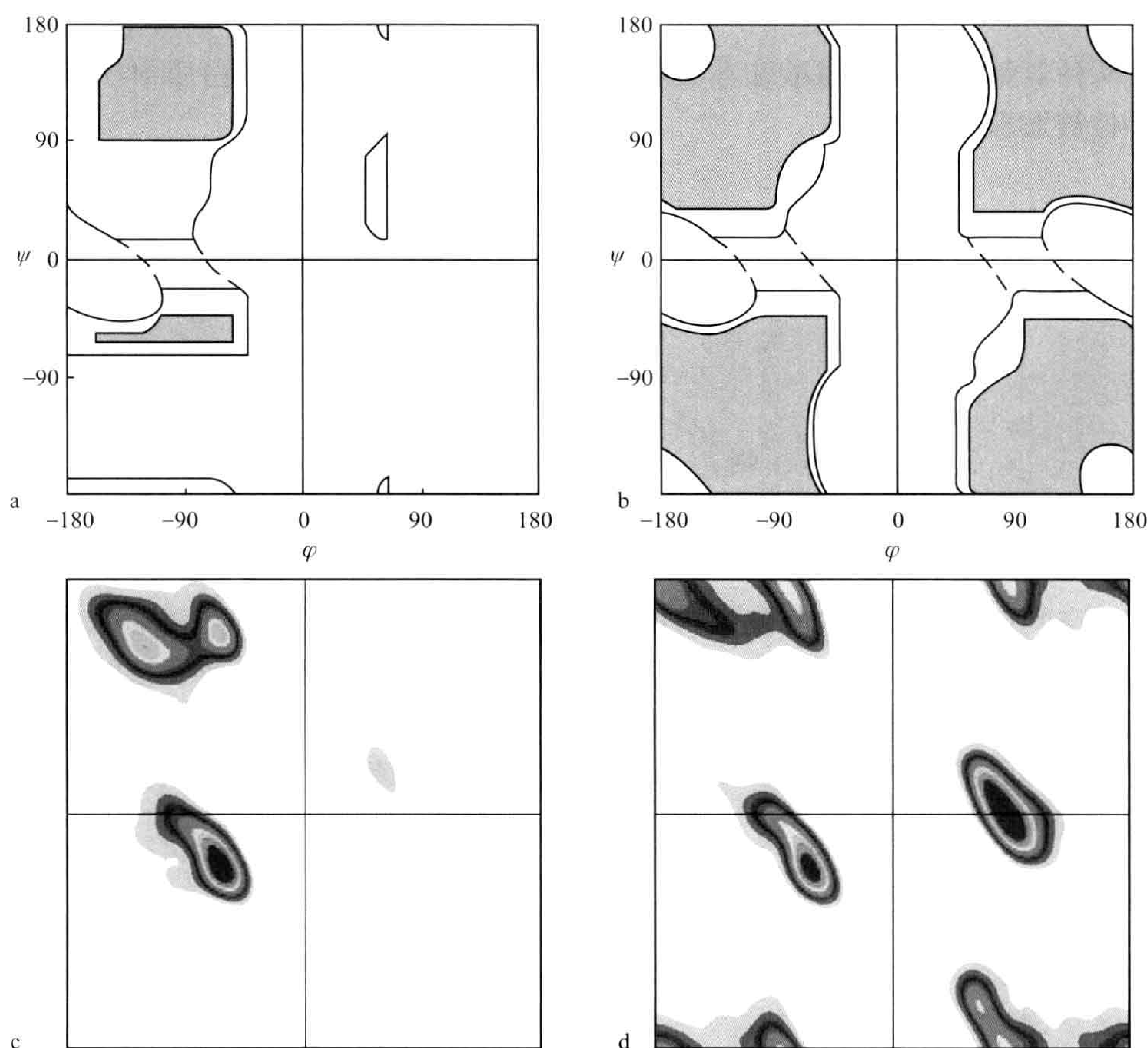


图 2.4　拉氏图（Ramachandran plot）。上：根据拉玛钱德安最初对丙氨酸（左）及甘氨酸（右）的空间位阻分析得到的允许区域；下：根据约 1000 个高分辨率蛋白质结构实际观察到的角度，非甘氨酸残基（左）和甘氨酸（右）。β 区被分为两块。注意实际观察到的分布非常不同于根据空间位阻所期待的分布，特别是对于甘氨酸残基 [得到许可重印自 Hovmöllers *et al*. (2002) Conformations of amino acids in proteins. [*Acta Cryst* **D58:** 768–776. Copyright (2003) Elsevier.]。

对很多结构确定得非常好的蛋白质构象角分布进行作图，β 区域清楚地被分为两块，α 区域也变成不同的形状（图 2.4c）。甘氨酸（Gly）残基的构象角分布图也显现出很不同的图案，在接近左手螺旋构象区域出现一个强峰（图 2.4d）。

拉氏图可被用来分析实验得到的蛋白质结构模型的总体质量。对于确定得很好的蛋白质结构，绝大部分构象角会处于被允许区域。蛋白质中的个别残基（构象角）仍然可能出现在不太被偏好甚至于不允许区域。在很多情况下，这些不被允许的构象是与该蛋白质的重要生物学性质相关联的。

2.1.3 侧链构象

有两种氨基酸，即亮氨酸和苏氨酸的 β- 碳是手性的。一般蛋白质中仅有 β- 碳的一种手性被发现（图 2.5）。

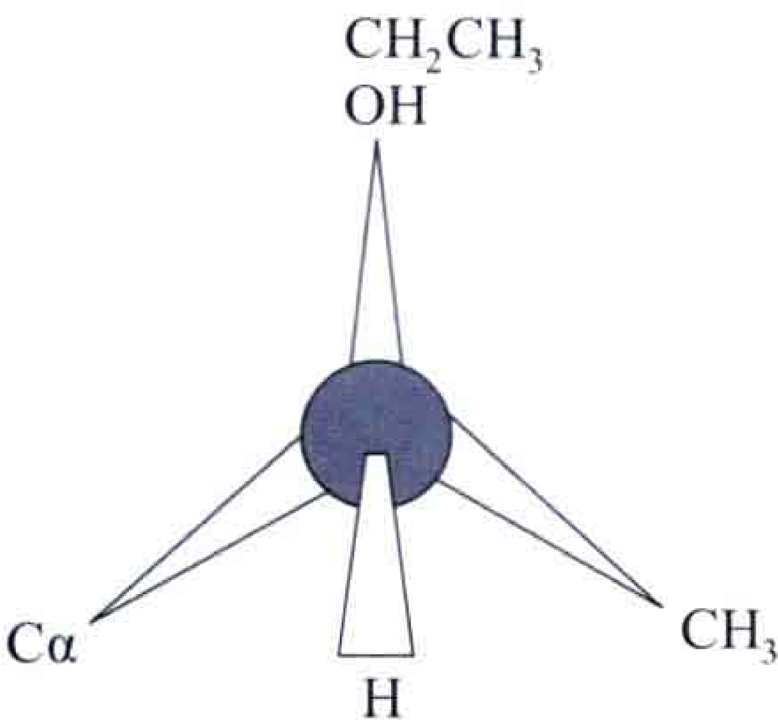


图 2.5 从 H-Cβ 键向下看 β - 碳原子时，蛋白质中异亮氨酸及苏氨酸常见构型的“CARC”法则。

由于空间位阻，侧链也有偏好构象。每个四面体碳原子的侧链有 4 种取代基，这些取代基偏好交错的构象且尽量远离其他原子（图 2.6）。这导致 3 种互换的 120° 间隔的扭转角。蛋白质结构分析表明氨基酸残基对于某些角度组合具有很强的偏好性。这些偏好构象，即旋转异构体（rotamer），可以收集成为数据库，对于蛋白质建模（modeling）非常重要（表 2.2）。

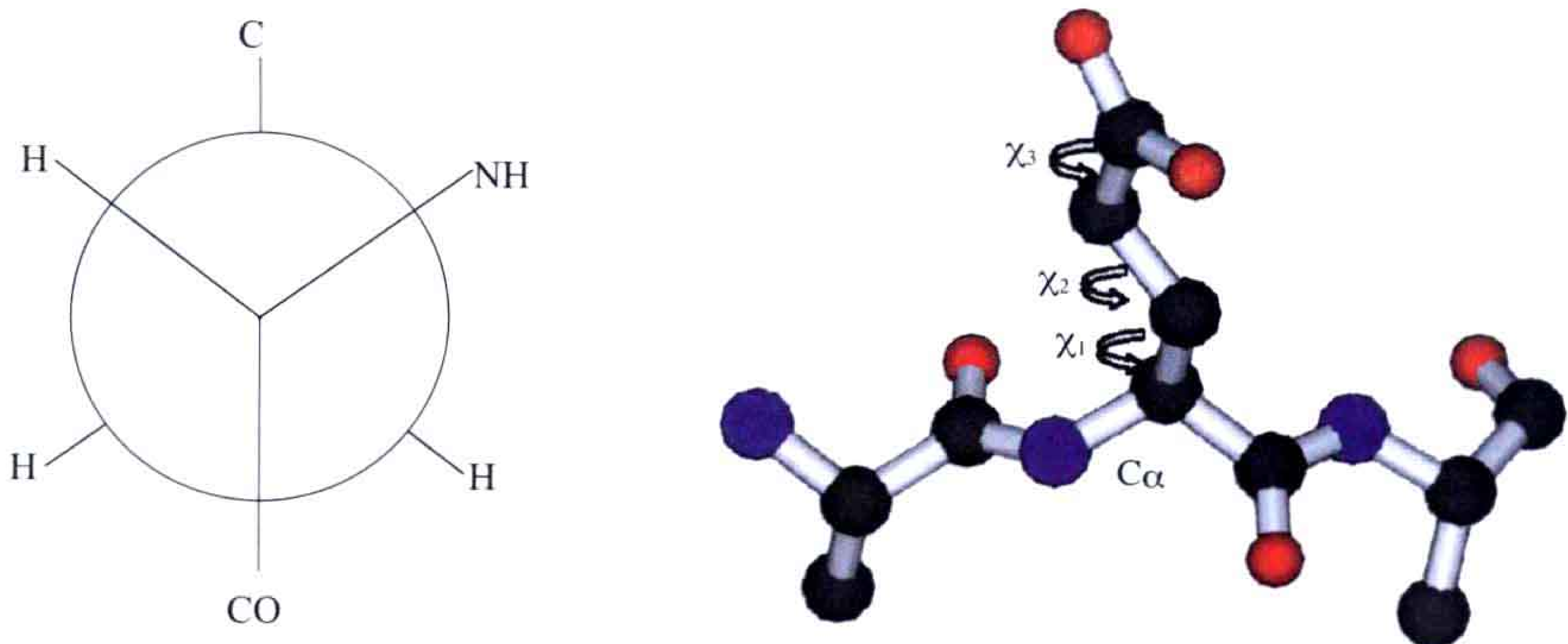


图 2.6 侧链中的扭转角图示。χ_1 角是 N-Cα 键和从 Cβ 到 Cγ（或者一个其他 γ 原子）之间的角度。左：顺着 Cα-Cβ 键看，χ_1 角为 –60 ° 时的交错构象。其他交错构象相应于 + 120 ° 或者 –120 ° 的旋转。右图显示谷氨酸（Glu）侧链在 χ_1 及 χ_2 接近 180 ° 时的常见旋转异构体。

表 2.2　4 种氨基酸残基中旋转异构体的偏好性

残基	旋转异构体	chi1	chi2	chi3	发生频率
GLU	–t0	–70	–177	–11	0.27
GLU	tt0	–176	175	–7	0.26
GLU	––0	–65	–69	–33	0.11
GLU	–+0	–56	77	25	0.09
GLU	+t0	70	–179	7	0.09
GLU	t+0	–174	71	14	0.06
GLU	+–0	63	–80	16	0.05
HIS	––	–63	–74	—	0.34
HIS	t–	–175	–88	—	0.25
HIS	–+	–70	96	—	0.16
HIS	+–	68	–81	—	0.14
HIS	t+	–177	101	—	0.09
HIS	++	48	86	—	0.02
ILE	–t	–61	169	—	0.45
ILE	––	–60	–64	—	0.18
ILE	+t	62	164	—	0.16
ILE	tt	–167	166	—	0.13
ILE	t+	–175	72	—	0.03
VAL	t	174	—	—	0.67
VAL	–	–63	—	—	0.26
VAL	+	69	—	—	0.05

注：缬氨酸仅有一个扭转角，组氨酸及异亮氨酸有两个，谷氨酸有三个。实际角度值是相应种类的旋转模观测值的平均。偏好的交错构象标记为 trans（180°）、gauche +（60°）及 gauche -（-60°，在上表中记为 t, + 和 –）。对于谷氨酸、谷氨酰胺及其他一些残基，观察到的旋转异构体构象与理想交错构象有偏差。在此表中记为“0”。摘自 Ponder JW, Richards FM. (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. (*J Mol Biol* 193: 775–791.)

2.1.4　二级结构

拉氏图显示蛋白质主链构象是受限的。一连串氨基酸常常采取相同的构象。这就导致了蛋白质的“二级结构”。蛋白质主链具有两种主要的规则二级结构，对应于拉氏图中的两块主要区域（图 2.4）。它们是 alpha(α) 螺旋及由几个 β- 束形成的 beta(β) 片层。α 螺旋与 β 片层二级结构在蛋白质内部能够有效地满足氢键的受体和供体的需要从而参与氢键形成（附录 A.3）。蛋白质结构中没有规则构象角的部分被称为环形或卷曲区域（loop or coil region）。它们常常出现在蛋白质的表面。这些区域有时会具有柔性，但这并非一定之规。

2.1.4.1 各种螺旋，特别是 α 螺旋

一条多肽链原则上可采取几种不同的螺旋构象。实践中只有三种有规律的构象出现（表 2.3）。最常见螺旋类型被称为 α 或 3.6_{13} 螺旋。在这个术语中，数字 3.6 表示每圈螺旋的残基数目，而下标（13）表示在这圈由主链 NH 和 CO 基团形成的氢键环中原子的数目。在 α 螺旋中，蛋白质主链在羰基的氧原子及下一个周期（残基 *n* 和 *n*+4）的 NH 的氮原子间形成氢键。每个残基有一个 1.5Å 的上升（图 2.7）。

每个侧链大约以 100° 的夹角指出螺旋轴，因而所有 Cβ 原子都指向螺旋的 N 端。

表 2.3 描述蛋白质中螺旋的一些参数

类型	3_{10}	α	π
每圈螺旋残基数	3.0	3.6	4.1
氢键原子数	10	13	16
氢键连接	*n* – *n*+3	*n* – *n*+4	*n* - *n*+5
相邻残基间的夹角	120	100	88
每个残基螺旋上升高度	2.0	1.5	1.15
ϕ /°	–75	–60	–75
ψ /°	–5	–45	–40

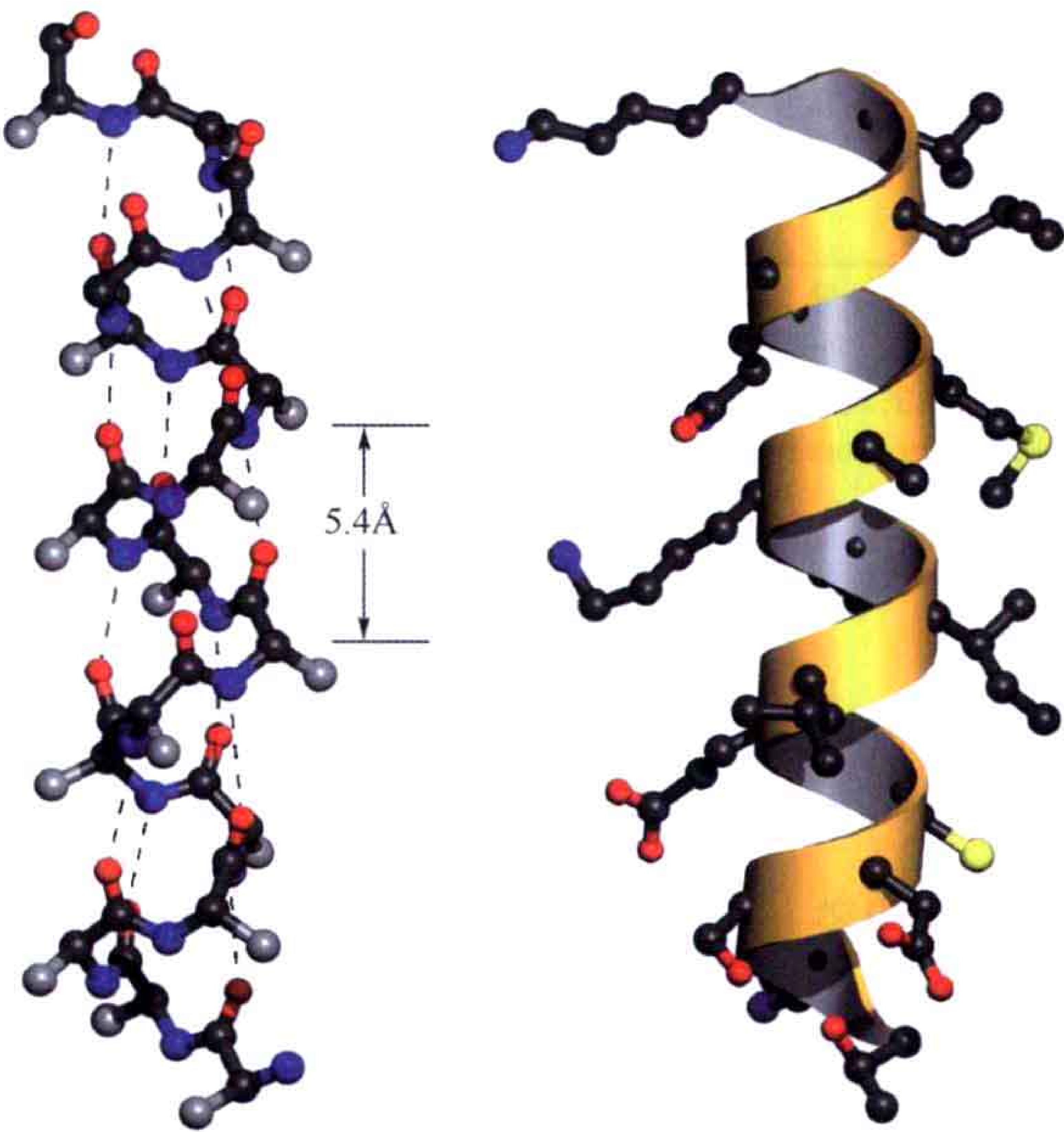


图 2.7 α 螺旋。左：一个典型 α 螺旋的主链及 C 原子（灰色）。其螺旋周期为 5.4Å。右：与左图同样的 α 螺旋显示出所有侧链。飘带显示主链，Cβ 原子指向螺旋的 N 端。

2.1.4.2　3_{10} 螺旋及 π 螺旋

除了 α 螺旋外，还有其他种类的螺旋。最常见的是 3_{10} 螺旋，其氢键是在残基 n 与残基 n+3 之间形成。3_{10} 螺旋比 α 螺旋窄一些，也不如 α 螺旋稳定，很少见长 3_{10} 螺旋存在的例子（图 2.8）。

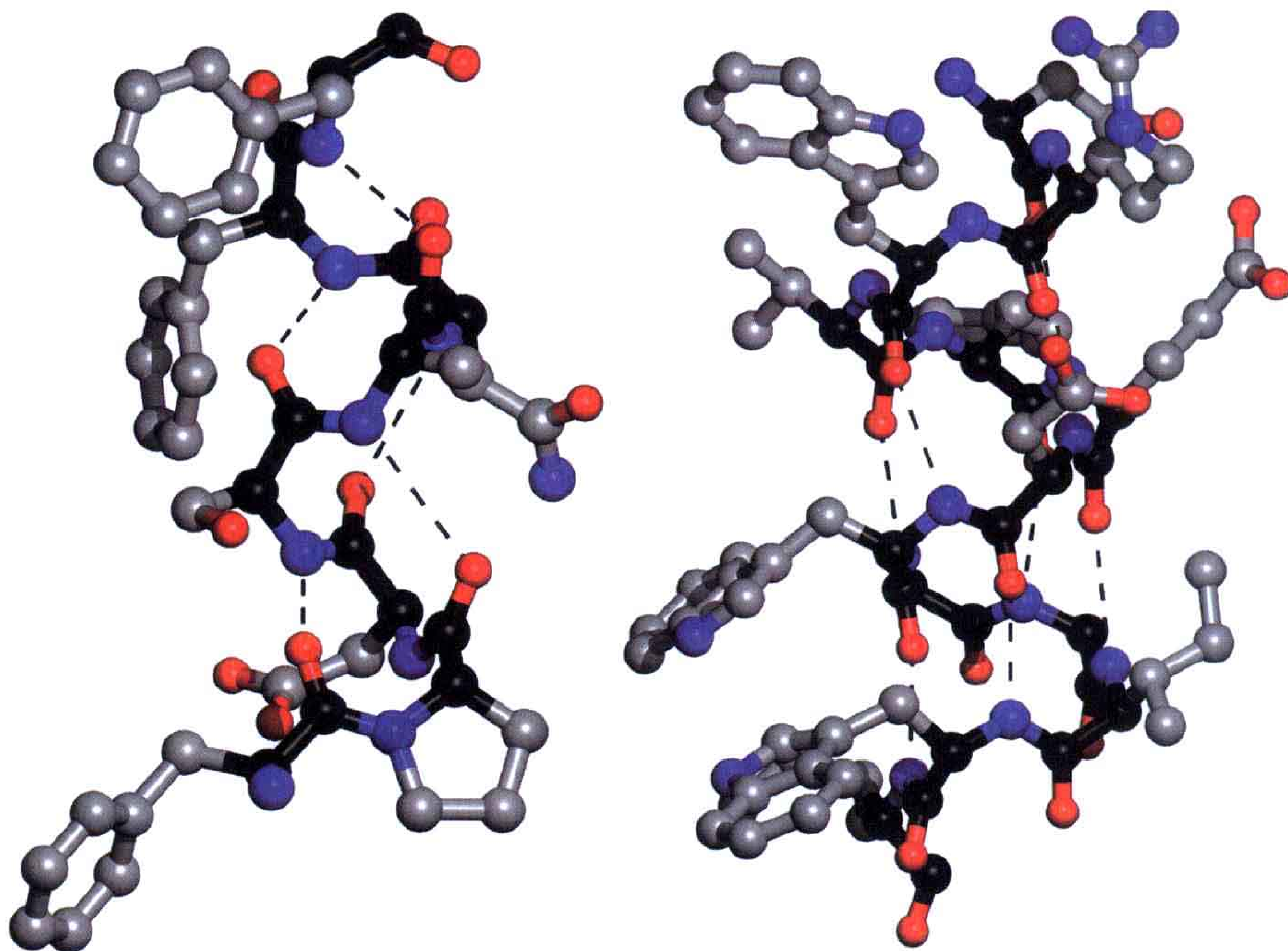


图 2.8　*Aplysia limacina* 肌红蛋白（PDB:1MBA）中的一个 3_{10} 螺旋及 *Methylococcus capsulatus* 的甲烷单加氧酶羟化酶 (PDB: 1MTY) 中的一个 π 螺旋。

另外还有一种螺旋被称为 π 螺旋。它的氢键是在残基 n 与残基 n+5 之间形成的。这是最少见的螺旋形式。这种螺旋曾被分类为 4.4_{16} 。然而，对 π 螺旋仔细分析表明每个螺旋周期的残基数大约为 4.1，因此 4.1_{16} 应该更加合适（表 2.3）。每周期 4.1 个残基可以沿着螺旋较好堆积（图 2.8）。π 螺旋曾被认为很不稳定，因为较大的外围可能使螺旋中间生成一个洞。实际分析 π 螺旋时发现，因为肽平面沿着螺旋的方向倾斜，沿着螺旋看并没有明显的空洞存在。

一般情况下，螺旋终点的几个残基可能采用 π 螺旋的氢键模式，然而对于几百个蛋白质的分析表明，至少可以发现 7 个连续残基形成的 π 螺旋。这三种螺旋的侧链具有不同的相对取向。因而，3_{10}、α 和 π 螺旋的氨基酸序列在蛋白质表面指向内部的疏水残基及指向外面的亲水残基之间具有不同的距离。

各种螺旋都常表现出不规则性。主链的氢键模式可能会沿着螺旋长度方向变化。α 螺旋的氢键是主导模式，然而此类螺旋可能在两端变得更宽或更窄，其氢键也变成 π 或者 3_{10} 螺旋的模式。

在蛋白质中左手 α 螺旋和 3_{10} 螺旋并不罕见，但一般都很短。

2.1.4.3 螺旋的偶极矩

螺旋都有一个重要的共同特点。因为螺旋中所有肽键平面的取向基本上都是相同的，并且可以被看做是小的偶极矩，这样螺旋就变成更大的偶极矩。实际上，螺旋的氨基端（N 端）具有部分正电荷而羧基端（C 端）具有部分负电荷。螺旋因此经常会被 N 端带负电的侧链及 C 端带正电的侧链稳定。同样地，结合带电的配体如辅酶或者底物可以被螺旋两端的部分电荷加强。由于静电相互作用，排列成反平行的相邻螺旋也会互相稳定。

2.1.4.4 β 结构

另一种常见的二级结构类型即 β 片层。这是由伸展的多肽链，即 β 束组成的，其中主链的 CO 和 NH 基团可以和两边的束形成氢键（附录 A.3）。依据 β 束的方向不同，β 片层可以是平行、反平行或者混合型的（图 2.9）。在一个理想的片层中，内部 β 束的所有 NH 基团都与 CO 基团形成氢键，在片层边上的 β 束会留下自由的（没有形成氢键的）NH 和 CO 基团。

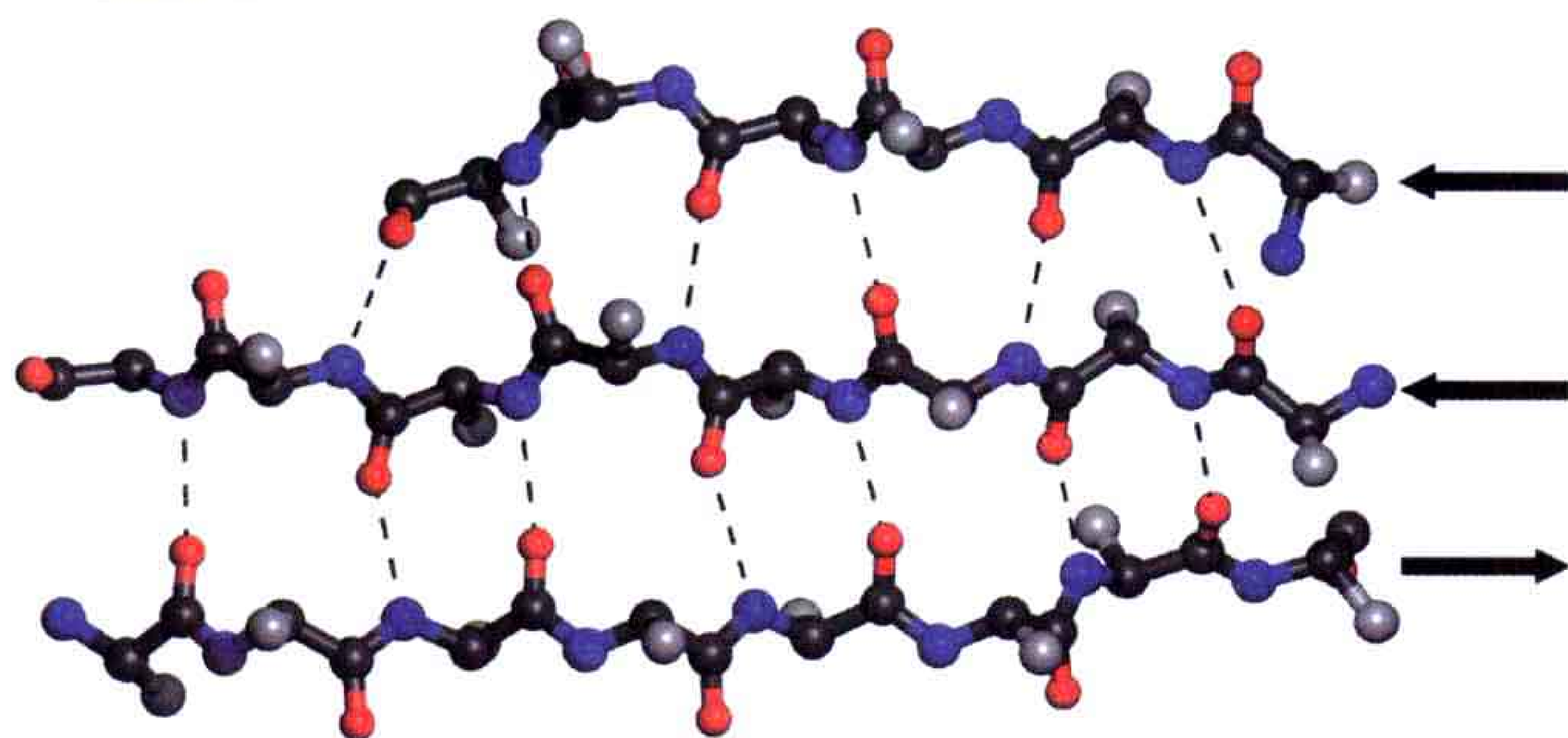


图 2.9 混合 β 片层中的氢键连接。β 片层总是有一些扭转。

在 β 片层中氨基酸侧链沿着主链（与片层垂直）交替伸展到片层的两侧。片层中相邻 β 束上的侧链指向相同方向——侧链的伸展方向与主链走向垂直。β 片层并非平坦，总是向着同一方向扭转（图 2.10）。如果顺着 β 束看，扭转是右手的。扭转的程度在

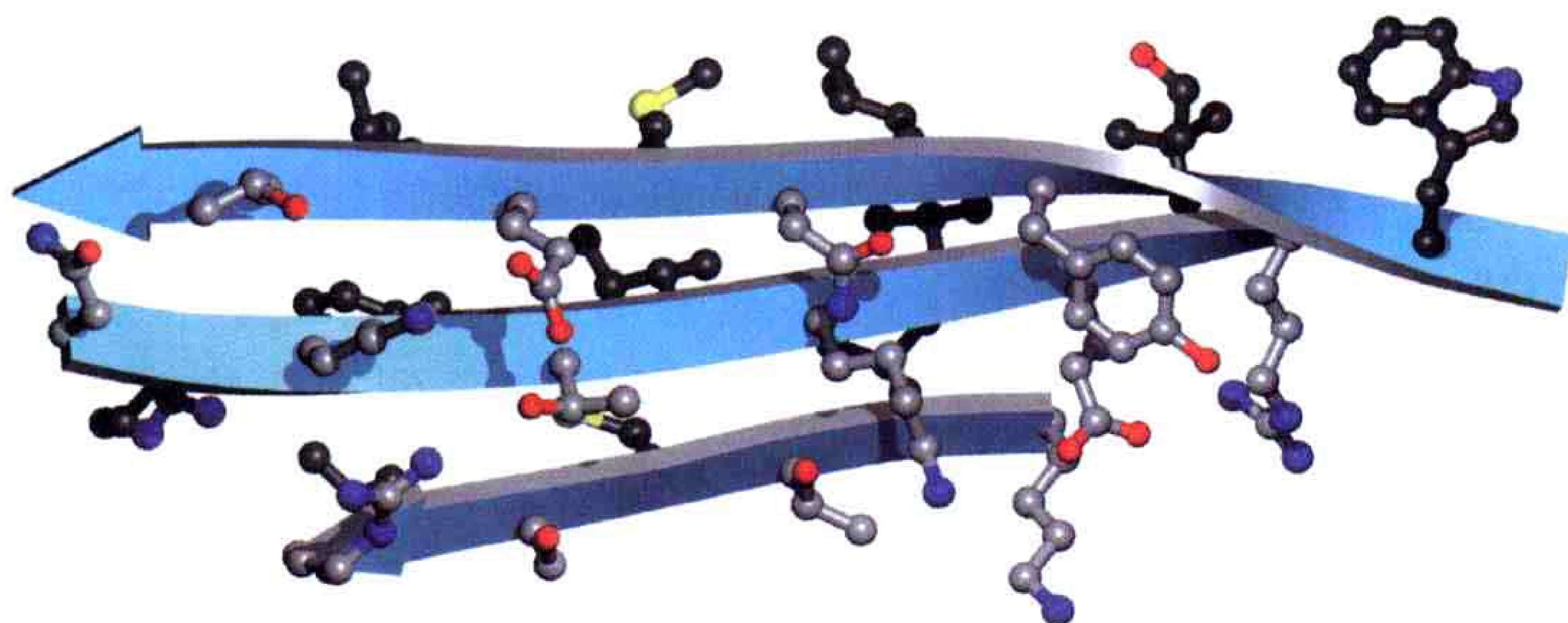


图 2.10 在一个反平行 β 片层中显示片层两侧的侧链位置。沿着 β 束看，侧链轮流指向片层两侧而 β 束间相邻的侧链则指向同一方向，并且可能相互作用。图中片层上面的侧链碳原子被涂成黑色，片层下面的侧链被涂成灰色（PDB:2BU1）。

不同片层间也是各不相同的。

片层边缘 β 束的一些 CO—及 NH—基团暴露于溶剂中具有与其他 β 束相互作用扩展片层的倾向。我们可以看到许多蛋白质中的亚基聚集就是以 β 片层的配对完成的。同样地，肽链底物结合蛋白酶是通过 β 束相互作用进行的。蛋白质的淀粉样聚集也是以某一部分间很强的 β 束相互作用导致的，这些部分在正常情况下会整合为折叠好的结构，从而在构象上与淀粉样聚集的形式很不同（2.2.3.4 节）。

2.1.4.5 β 突起

在反平行的 β 结构中，经常会看到突起。这通常发生在某一 β 束中多出一个氨基酸残基的情形，与理想状况氢键模式（图 2.11）相比，这时残基的排列发生错位偏移。β 突起更增加了 β 片层的扭转。β 突起具有典型的标准构象，可以被包括在规则的二级结构中。

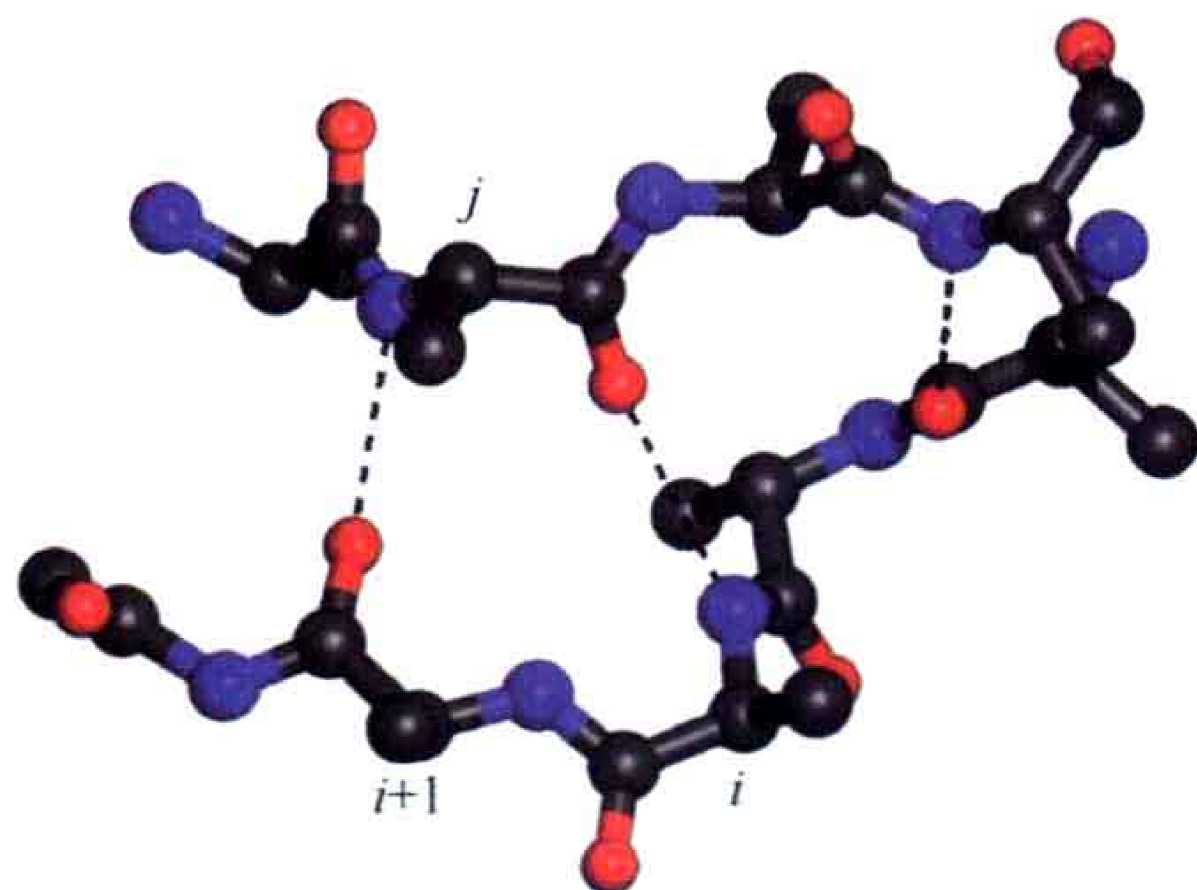


图 2.11 一个 β 突起。上链残基 *j* 的 NH 和 CO 基团分别与下链残基（*i*）的 NH 基团及残基（*i*+1）的 CO 基团形成氢键。

2.1.4.6 聚脯氨酸螺旋及胶原蛋白

富含脯氨酸的序列会形成特殊类型的二级结构，即多脯氨酸型二级结构。在拉氏图中，多脯氨酸构象位于 β 结构区域的 $\phi = -60^{\circ}$ 及 $\psi = 160^{\circ}$。这类二级结构是胶原蛋白中的主要结构，胶原蛋白由三股高度重复的序列（脯氨酸 - 羟脯氨酸 - 甘氨酸）$_n$ 构成，多脯氨酸链互相缠绕形成特殊类型的右手三股螺旋（图 2.12；13.1 节）。这类结构有时也罕见于水溶性蛋白如补体系统的 C1q 中。

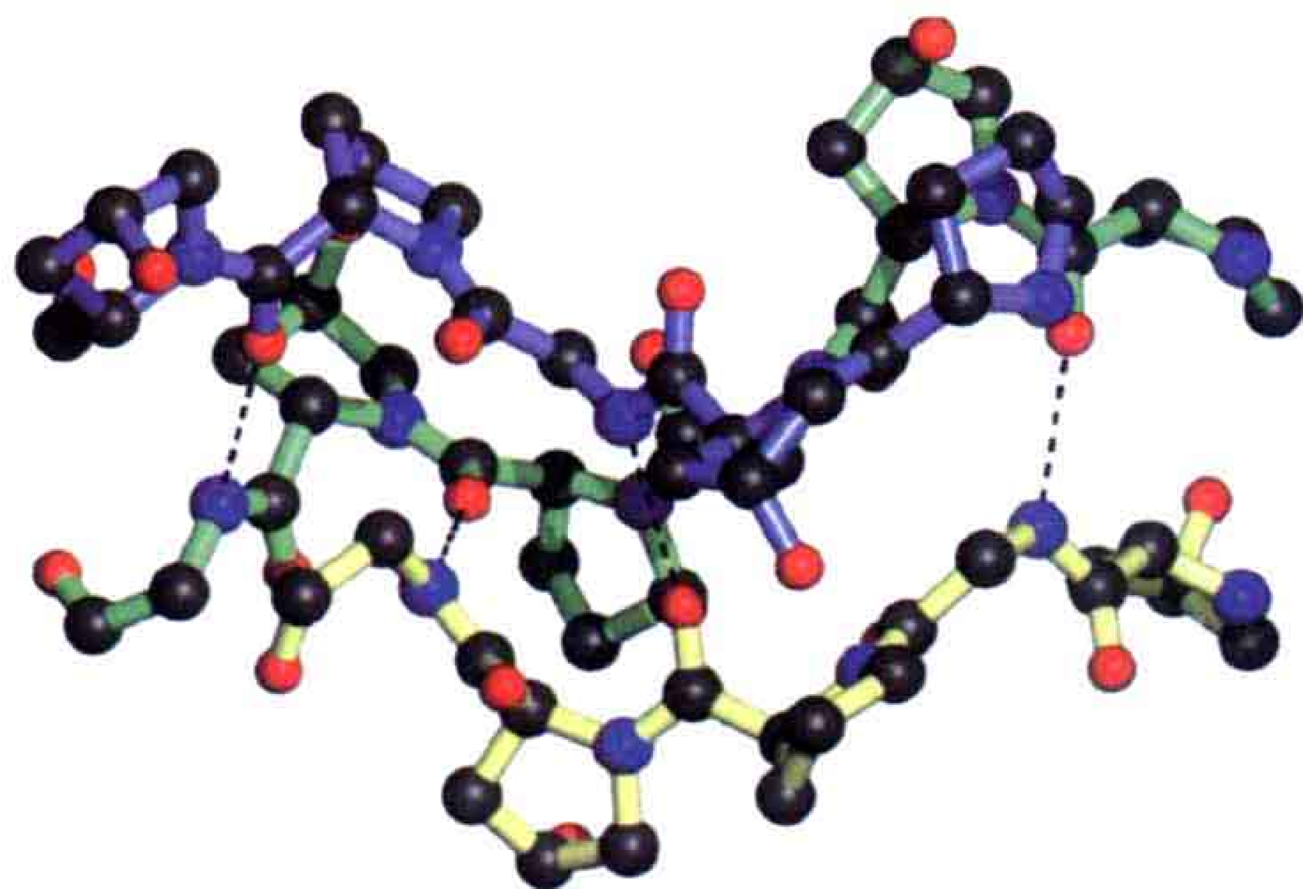


图 2.12 胶原蛋白中的一段三重螺旋。三条链中都有重复序列脯氨酸 - 羟脯氨酸 - 甘氨酸。甘氨酸残基允许三条链密切接触形成链间氢键 (PDB:1Q7D)。

2.1.4.7 各种转角

反转角是另一种类型的规则二级结构，其短小且经常连接两条 β 束。反转角或 β 转角在理想情况下由残基 *n* 的 CO 基团与残基 *n*+3 的 NH 基团形成氢键（图 2.13）。这类紧致的转角对于残基 *n*+1 和 *n*+2 附加了很强的构象限制（表 2.4）。蛋白质存在好几类这样的转角，其中一些对于特定残基（Gly、Pro，它们的化学性质对于这些转角的稳定很重要）在特定的位置具有很强的偏好性。

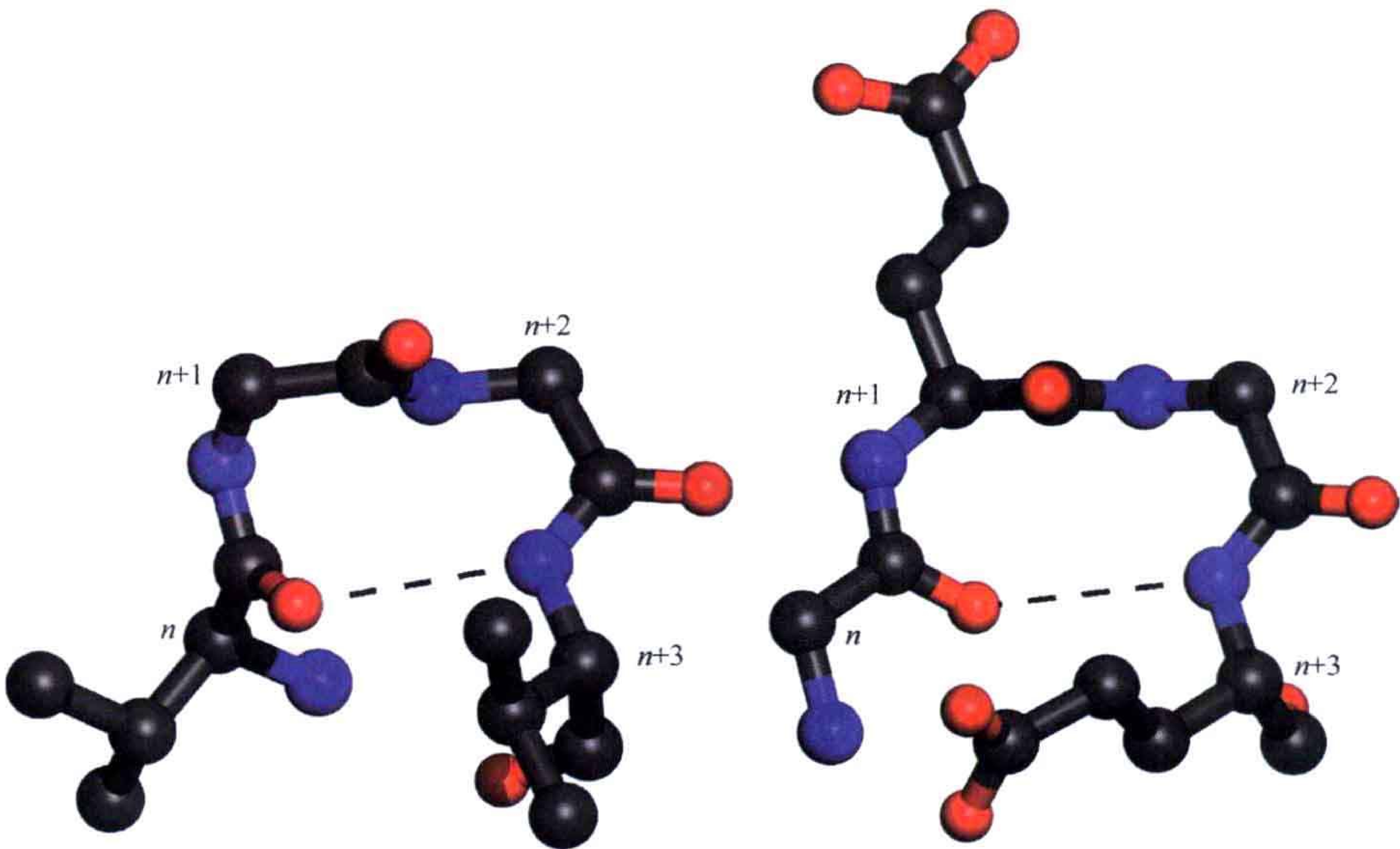


图 2.13 两种最为普通的反转角，I′ 和 II 型反转角。

表 2.4 一些最为常见的反转角或 β 转角的构象角及氨基酸残基偏好性

类型	Phi	Psi	Phi	Psi	氨基酸偏好性
I	–60	–30	–90	0	Pro(*n*+1), Gly(*n*+3)
I′	60	30	90	0	Gly(*n*+2)
II	–60	120	80	0	Pro(*n*+1), Gly(*n*+2),
II′	60	–120	–80	0	Gly(*n*+1)
VIa	–60	120	–90	0	
VIb	–135	135	–75	0	
VIII	–60	–30	–120	120	

注：*n* 是转角的第一个残基。构象角是由残基 *n*+1 和 *n*+2 构成。

另一种许多蛋白质中都具有的转角是 γ 转角，其中残基 *n* 的 CO 基团与残基 *n*+2 的 NH 基团间形成氢键。

2.1.5　拓扑及模体

2.1.5.1　反平行 β 片层

二级结构元素及这些二级结构间的连接构成蛋白质的拓扑（结构）及蛋白质的三级结构。对已知蛋白质的分析表明有些模块比其他一些常见得多。例如，β 片层常由 β 发夹构成。如果 β 发夹模块重复，则形成“上下片层”，有时也称 β 回纹波形（图 2.14）。

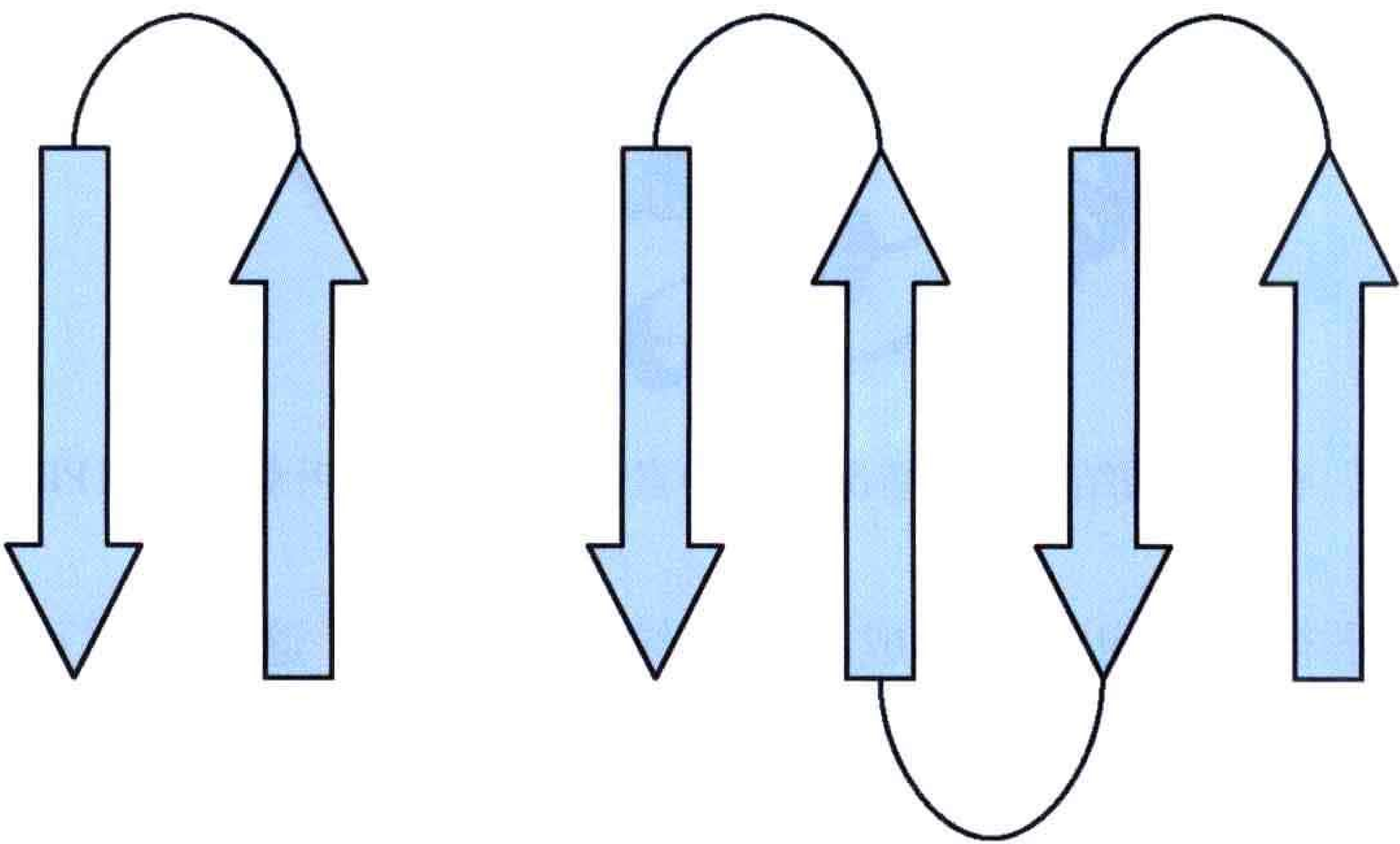


图 2.14　一个 β 发夹及其构成的四束“上下片层”结构。

“上下片层”在许多蛋白质中都存在。它们经常形成一大片开放片层，但有时也会形成桶形或圆柱状（图 2.15）。有一大类蛋白质被称为“螺旋桨”蛋白，其每个桨片由 4 束反平行上下片层构成（图 2.16）。四片、五片、六片、七片、八片及十片“螺旋桨”蛋白都有很多。大多数这类蛋白质中，其中一个桨片是由 N 端一束及 C 端三束 β 片层组成，这样利于封闭及锁住螺旋桨结构。

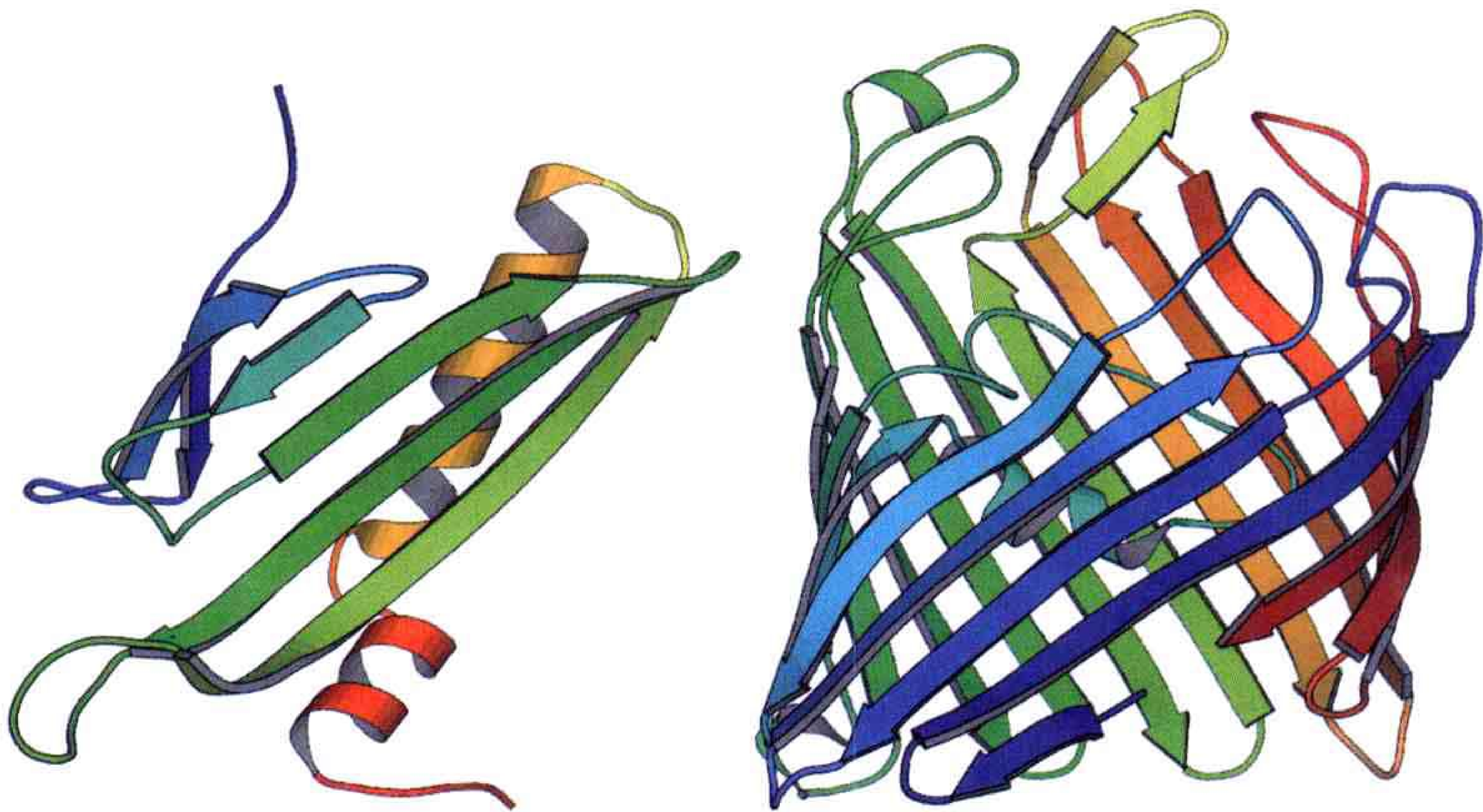


图 2.15　两个“上下片层”。噬菌体 MS2 包被蛋白的开放片层（PDB:2MS2），以及封闭柱形的细菌膜孔蛋白——大肠杆菌外膜上的一个膜孔蛋白（PDB:2OMF）。

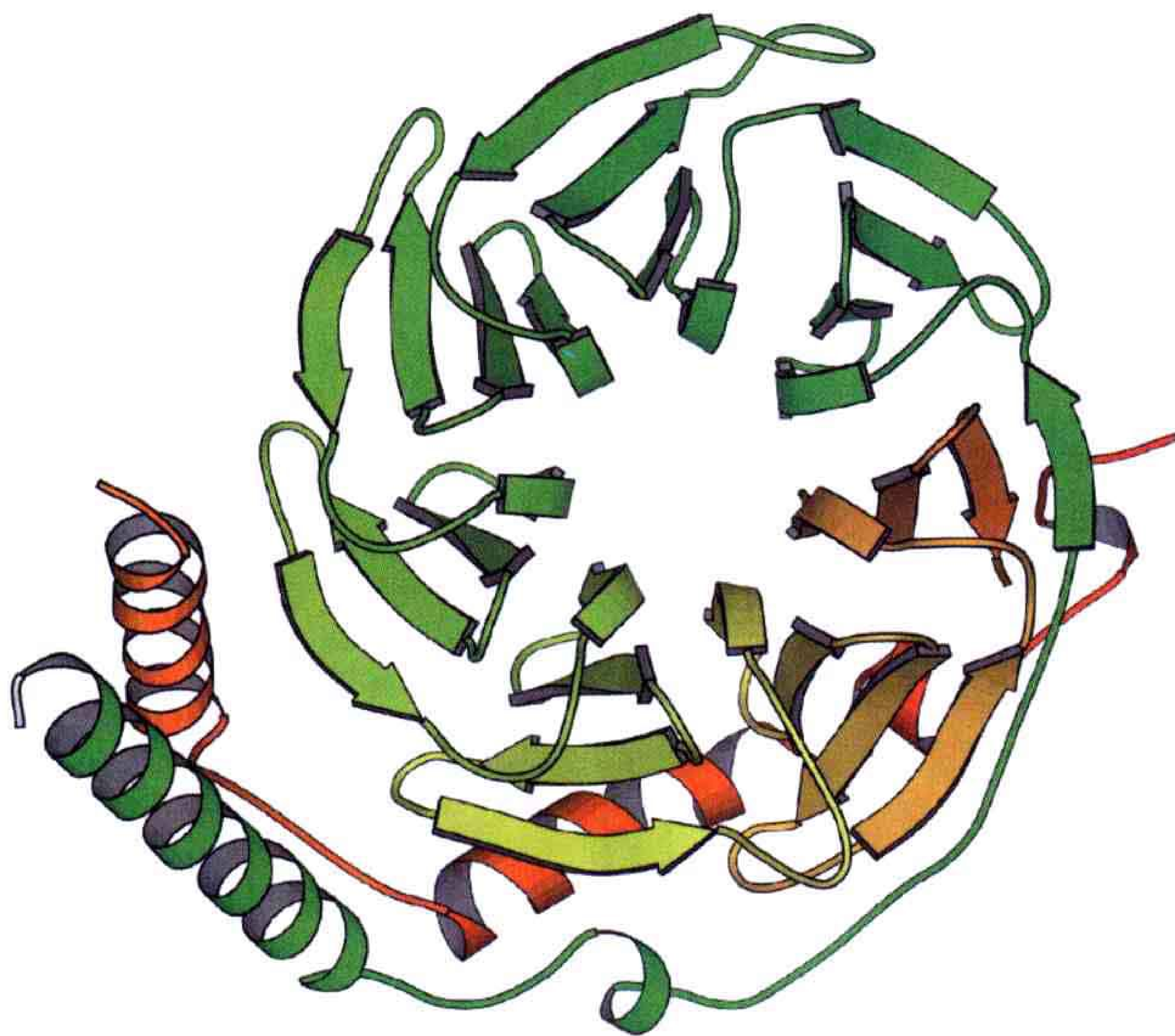


图 2.16 “上下片层”在螺旋桨结构中的情况。传导素 β 亚基中的七叶螺旋桨（PDB:1GOT）。

一种四束反平行片层也能够形成希腊回纹拓扑结构。这种模体可被看做是一个 β 发夹再被弯折形成 4 束。希腊回纹模体在 β 片层蛋白中也是非常常见的（图 2.17）。

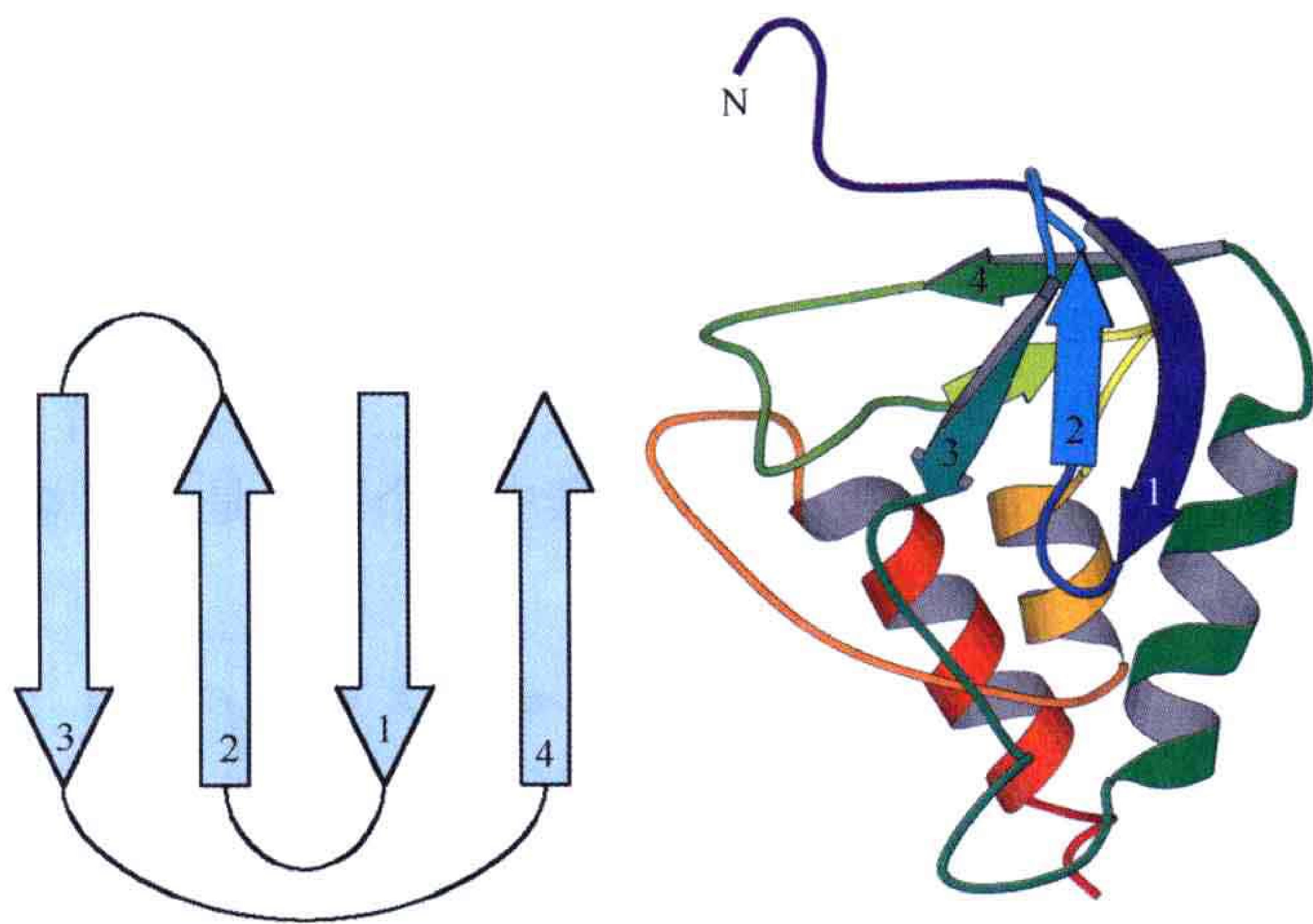


图 2.17 希腊回纹模体示意以及此模体在蛋白质结构中的实际情况 (Micrococcal nuclease, PDB: 2SNS) 5- 链片层和位于连接着链 3 和 4 的环状区域的螺旋组成一个在许多蛋白质中都存在的 OB(寡聚核苷酸及低聚糖) 折叠。

一种特殊的希腊回纹模体延伸出的结构被称为果冻卷（jellyroll）。这是一种由两个四束片层形成的八束结构（图 2.18）。

主要由 β 结构组成的蛋白质一般形成反平行片层，但有一个例外就是 β 螺旋拓扑结构，其肽链以短 β 束的形式构成三棱柱形的螺旋结构（图 2.19）。

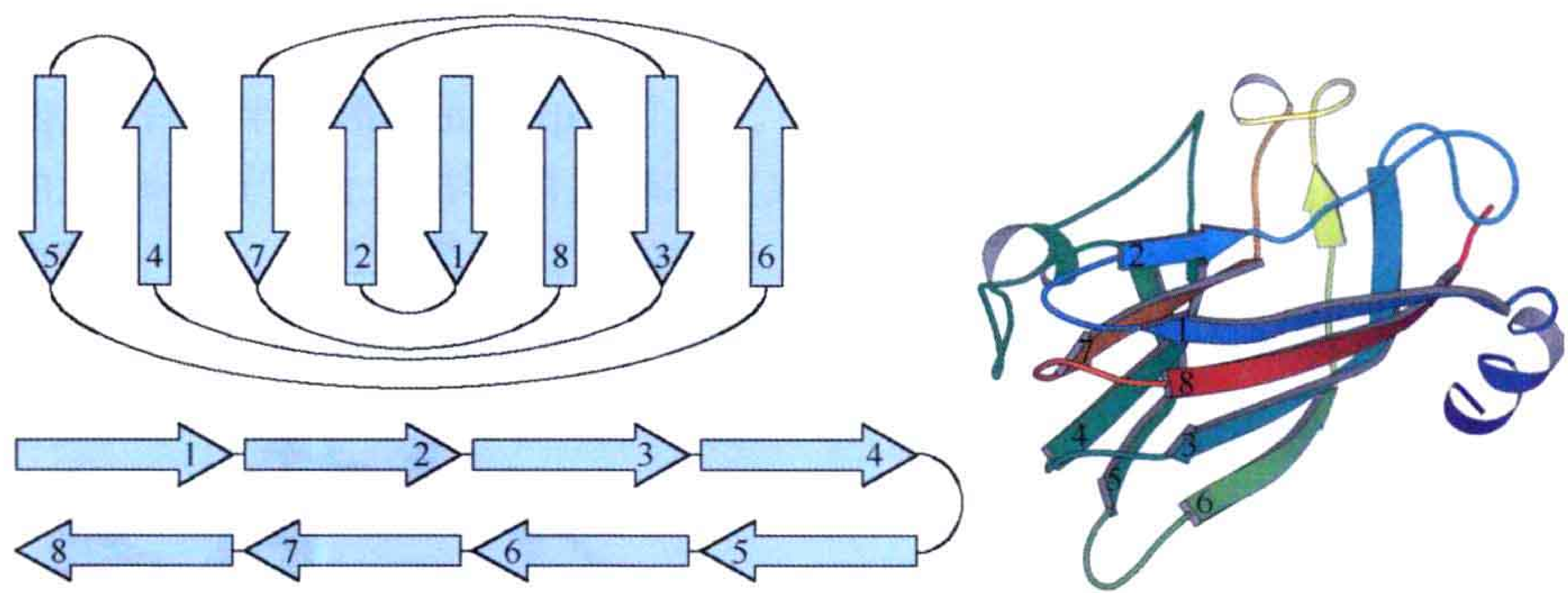


图 2.18 左：果冻卷拓扑结构（上）可以看做由 β 发夹（下）按照希腊回纹卷成的；右：植物病毒 STNV 的外壳蛋白，一个简单的果冻卷折叠（PDB:2BUK）。

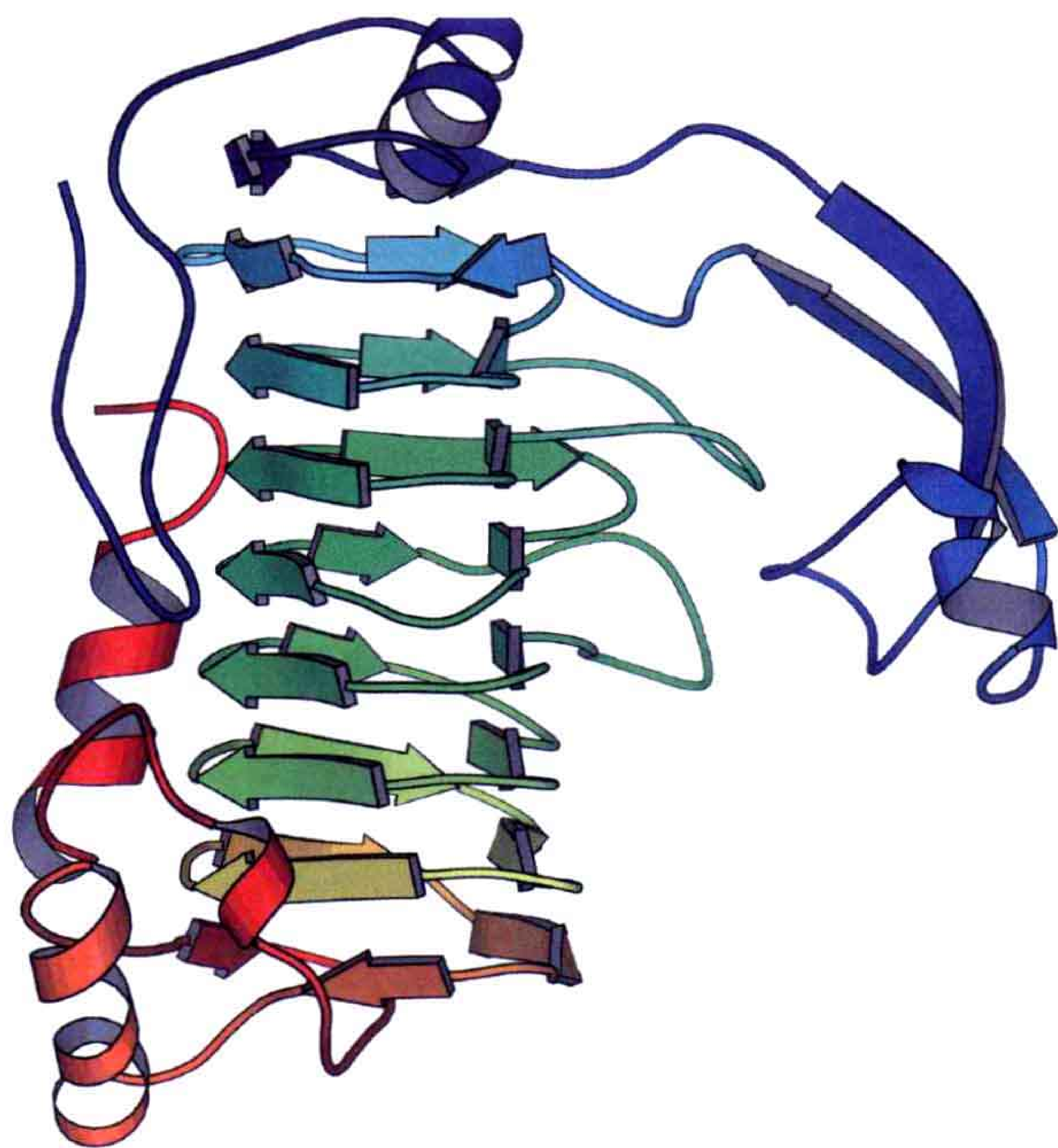


图 2.19 一个 β 螺旋蛋白：*Aspergillus niger* 的果胶裂解酶 (PDB: 1IDK)。

2.1.5.2 βαβ 单元

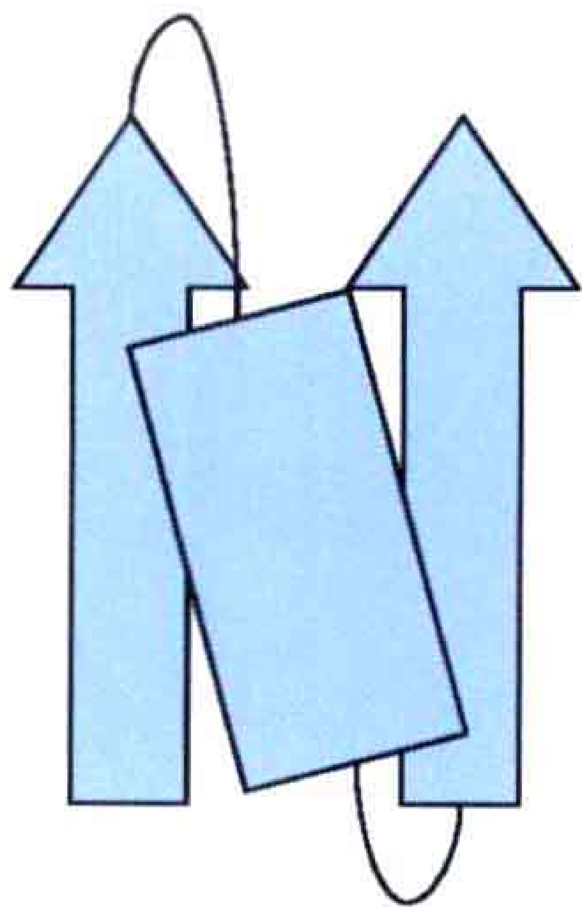


图 2.20 一个 βαβ 单元的示意图。沿着多肽链走向是右手旋进的。

蛋白质中另一个常见模体是 βαβ 单元，其由两个平行束及它们之间的连接螺旋构成。这个模体可以有两种形式存在，其中右手旋进版本是最为常见的（图 2.20）。

蛋白质中的平行片层多由 βαβ 单元构建而成。在这些蛋白质中，βαβ 单元多次重复构成一部分或者所有的片层结构。存在于许多蛋白质中的罗斯曼（Rossman）折叠，即由 βαβ 单元构筑而成（图 2.21）。罗斯曼折叠首先是在乳酸、马来酸及乙醇脱氢酶中被发现，其中

六束平行片层构成这些蛋白质的结构核心。

在罗斯曼折叠中，N端β束位于片层正中。两个βαβ单元各形成半个片层(βαβαβ)，其余结构是将类似的单元由一个大致的二次对称轴关联形成。由于这些单元的手性及二次对称关系，几个α螺旋会分别位于中心片层的两边（图 2.21）。许多蛋白质具有这样类似的拓扑结构，在β束的数目和秩序上有时会有细微变化。

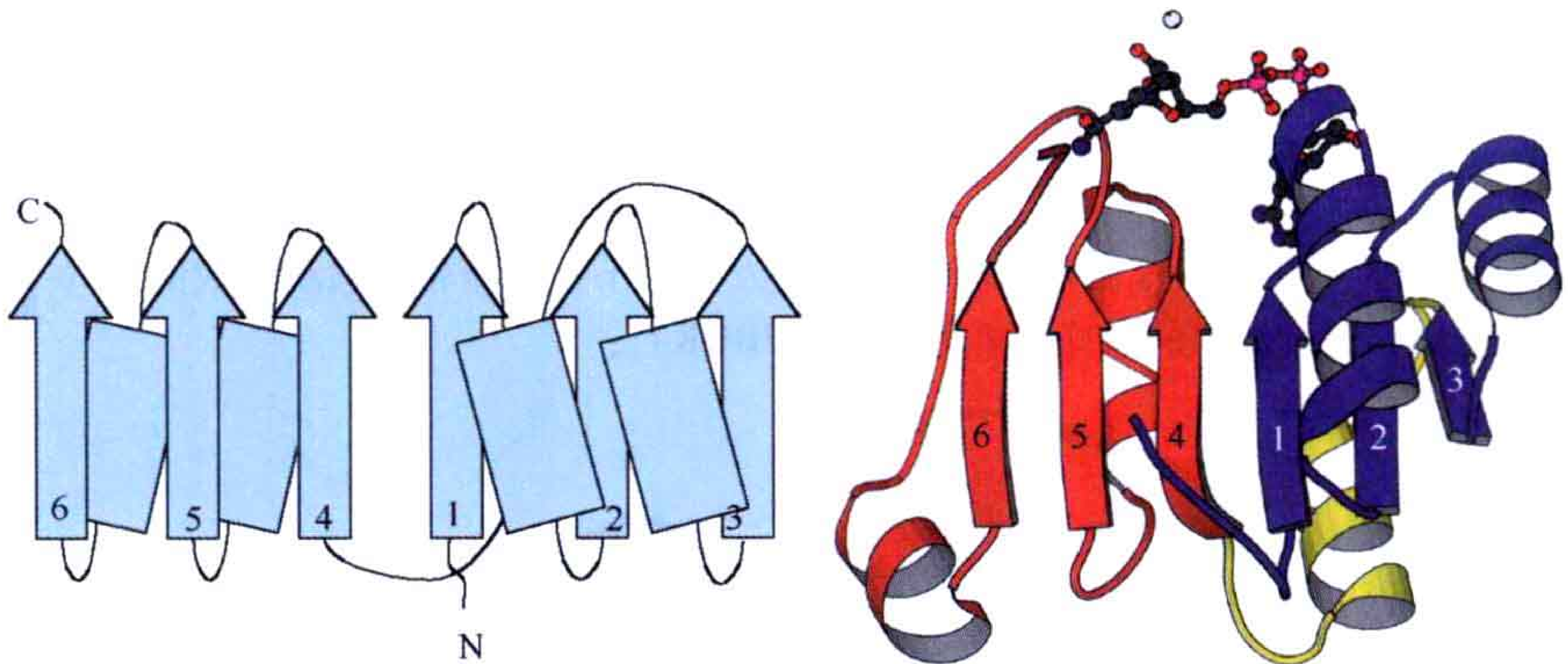


图 2.21　左：图示罗斯曼（Rossmann）折叠的拓扑结构；右：*Sulfolobus solfataricus* 乙醇脱氢酶的 NAD 结合结构域，一个典型的罗斯曼折叠（654123 的β束顺序）。结构域的第一个βαβαβ花样表示为蓝色，第二个为红色，连接部分为黄色。核苷酸结合在中心β片层的C端尾部。NAD分子的磷酸部分位于两个螺旋的N端，磷酸由于螺旋二极矩易于结合在此处 (PDB:1R37)。

另外一个很常见的折叠类型是 TIM 桶，其名字来源于第一次观察到此折叠的酶，即磷酸甘油醛异构酶 (triose phosphate isomerase)。这种折叠是由8个β–α单元形成一个圆筒（图 2.22）。这8个平行β束形成一个封闭的圆筒，8个连接螺旋包围在这个β圆筒外面。这种折叠存在于大量的酶中（以及一个非酶蛋白，narbonin），很多这类蛋白质并没有序列及功能相关性。

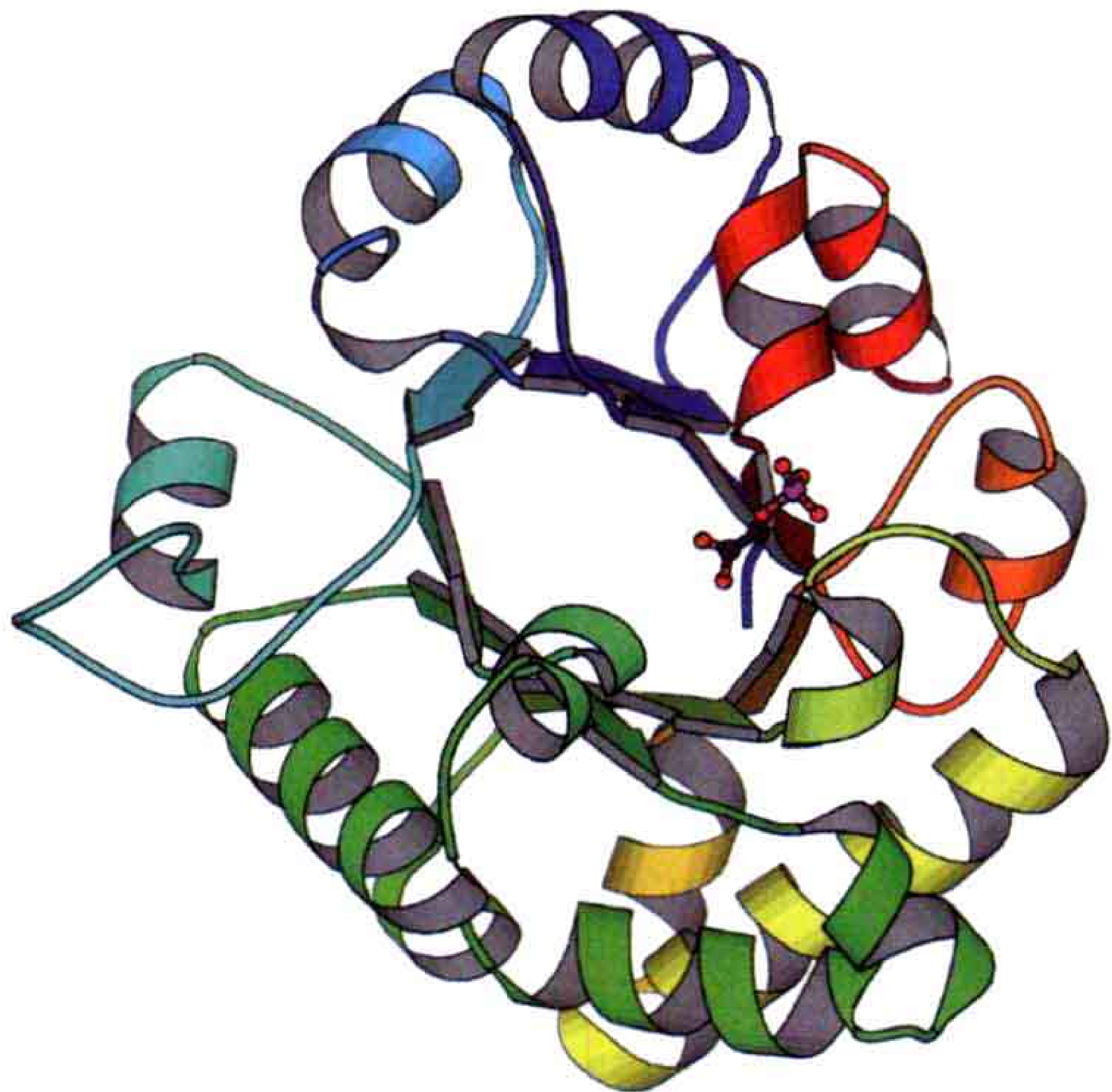


图 2.22　*Plasmodium falciparum* 磷酸甘油醛异构酶（TIM）的 TIM 桶结构。β束顺序为 12345678。活性中心被一个过渡态类似物——磷酸乙醇酸盐占据。在所有 TIM 桶结构中，β束C端部分总是位于活性中心 (PDB:1LYX)。

2.1.5.3　螺旋的堆积

富含 α 螺旋的蛋白质只有有限的几种方式来堆积这些螺旋。这是由于螺旋的侧链会形成各种“嵴”，最佳的堆积需要将一个螺旋的“嵴”置于其他螺旋“嵴”之间的沟槽中。这就导致了螺旋轴之间的角度为 20° 或者 50°。一种常见的螺旋堆积类型是“四螺旋束”，两对反平行的螺旋轴之间形成大约 20° 的角度堆积。四螺旋束存在于许多蛋白质中，一个简单的例子为，上下四螺旋束桶如图 2.23 所示，还有许多其他方式的螺旋束由 1、2 或者 4 个单元组成。球蛋白折叠是另一类主要的螺旋堆积类型（图 2.24）。

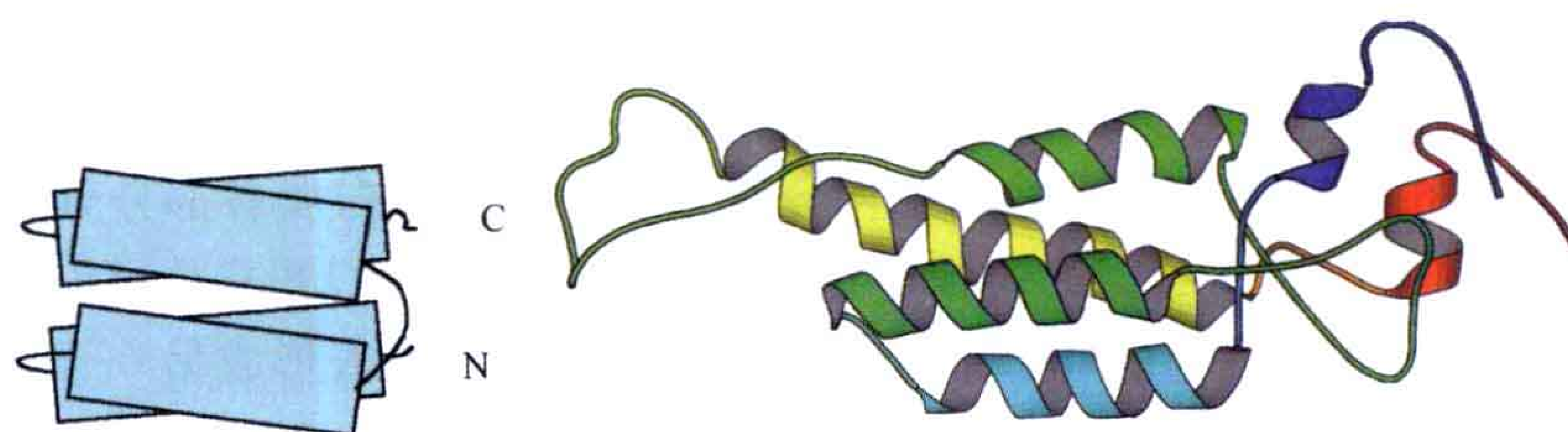


图 2.23　上下四螺旋束拓扑结构及烟草花叶病毒 TMV 的包被蛋白 (PDB:2TMV)。

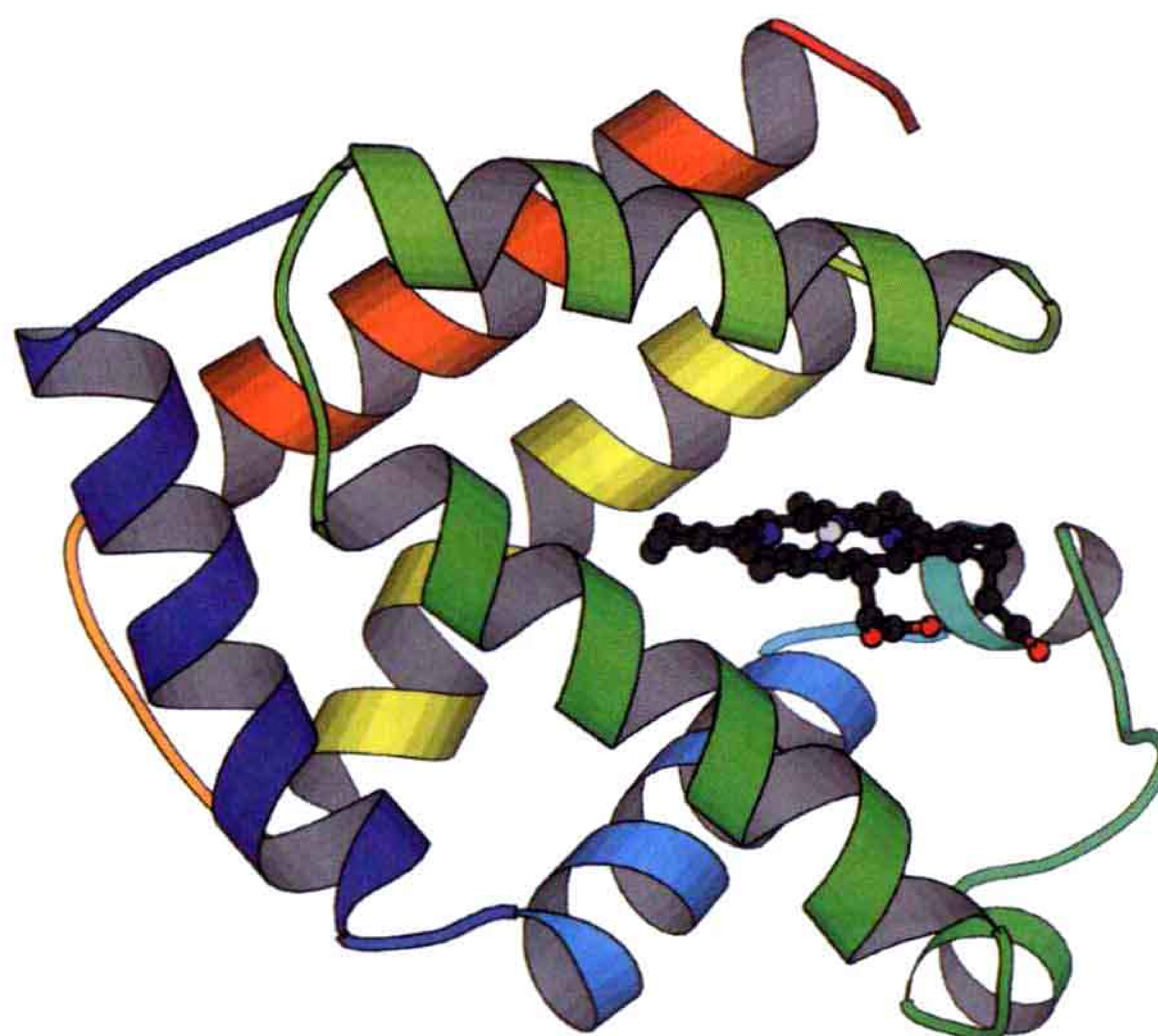


图 2.24　肌红蛋白的结构。其大部分螺旋堆积为螺旋轴，大约呈 50° 角。此蛋白质结合一个血红素辅基（PDB:1MBA）。

2.1.6　球形及纤维状蛋白质

蛋白质可以多种形状存在，其经典图像是球状结构（图 2.24）。最早解析出的蛋白质结构是肌红蛋白、血红蛋白及鸡蛋清溶菌酶，它们都呈现球状结构。然而，很久以来蛋白质的延伸或者纤维状结构也是已知的。头发中的角质蛋白由以 α 螺旋为主的纤维状蛋白质构成，蚕丝则由主要为 β 束的纤维状蛋白质构成。

2.1.6.1 结构域

许多蛋白质是围绕一个单一的疏水核心构成的，但是大多数蛋白质结构是由分开的折叠单元——结构域构成的。一个结构域通常是围绕单一疏水核心构成的蛋白质结构，这个描述可以作为结构域的定义。通常结构域是沿着多肽链排列的分离单元。蛋白质中一类结构域的形成是通过基因的复制与融合。同样的蛋白质折叠可以这种方式沿着多肽链重复一次或者多次。这里的一个经典例子是酸性蛋白酶，这种酶有时以两个完全相同的亚基以二重轴相关联构成。但是，在其他一些情况下这种酶会以一条多肽链上的两个结构域组成，结构域同源并与一个近似的二重轴相关联。这样的结构域组成有时可以利用 Cα-Cα 距离绘图显示出来（附录 B.1）。

很多情况下，蛋白质各部分之间的结构分解不明显。将蛋白质各部分分为结构域或者亚结构域则变得主观。尽管结构域是分立的折叠单元，它们并非总是沿着一条多肽链连续。有时也会有多肽链的几个部分从一个结构域延伸出去形成另一个独立的结构域，在这种情况下，结构域之间至少会有两个连接。例如，图 2.25 中红色结构域（螺旋结构域）插入了 Ras 结构域。Ras 结构域或称 G- 结构域（见 5.3.3 节和第十一章）在许多蛋白质中都存在并且以不同方式与其他结构域组合。G- 结构域结合并且水解 GTP。很可能 G- 结构域是一个古老的单元，被复制了多次并且演化出了许多不同的功能。要进行很多不同功能，与其他结构域的组合是关键。有时这些结构域插入 G- 结构域中间，但其他情况下，G- 结构域可以在其他结构域的 N 端或者 C 端延伸出来。

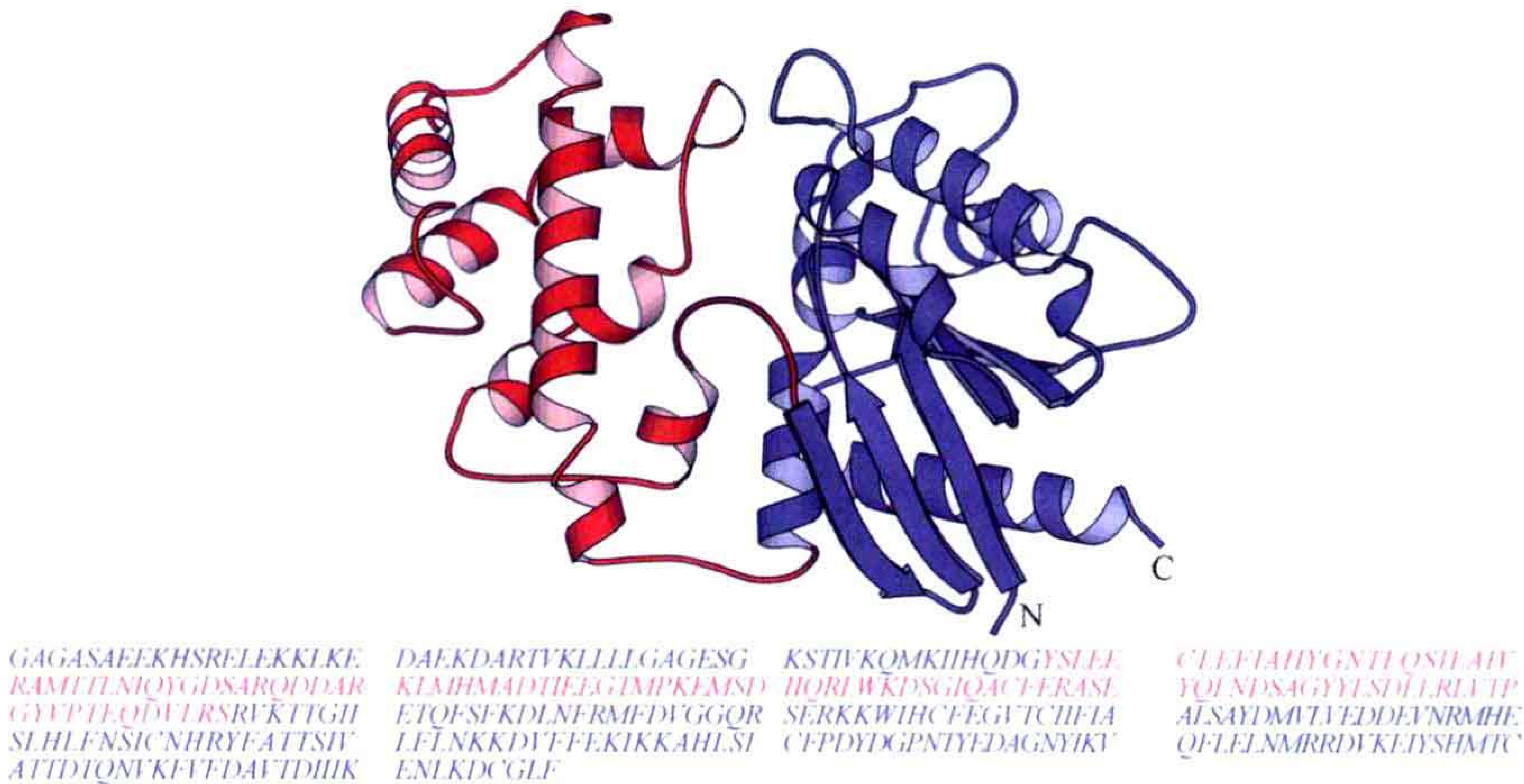


图 2.25　一个两结构域蛋白——传导素 α。一个结构域（红色）插入另一个 G- 结构域（蓝色）之中。下面的氨基酸序列以蓝色和红色相应地显示出主要及插入结构域。

2.1.6.2 结构域互换

包含几个结构域的单体蛋白有时可能以结构域互换的非天然方式聚集成为聚体。本该在同一条多肽链相互作用的结构域（或者其他二级结构元素）会互换而作用于另

一条多肽链的相应结构域（图 2.26）。

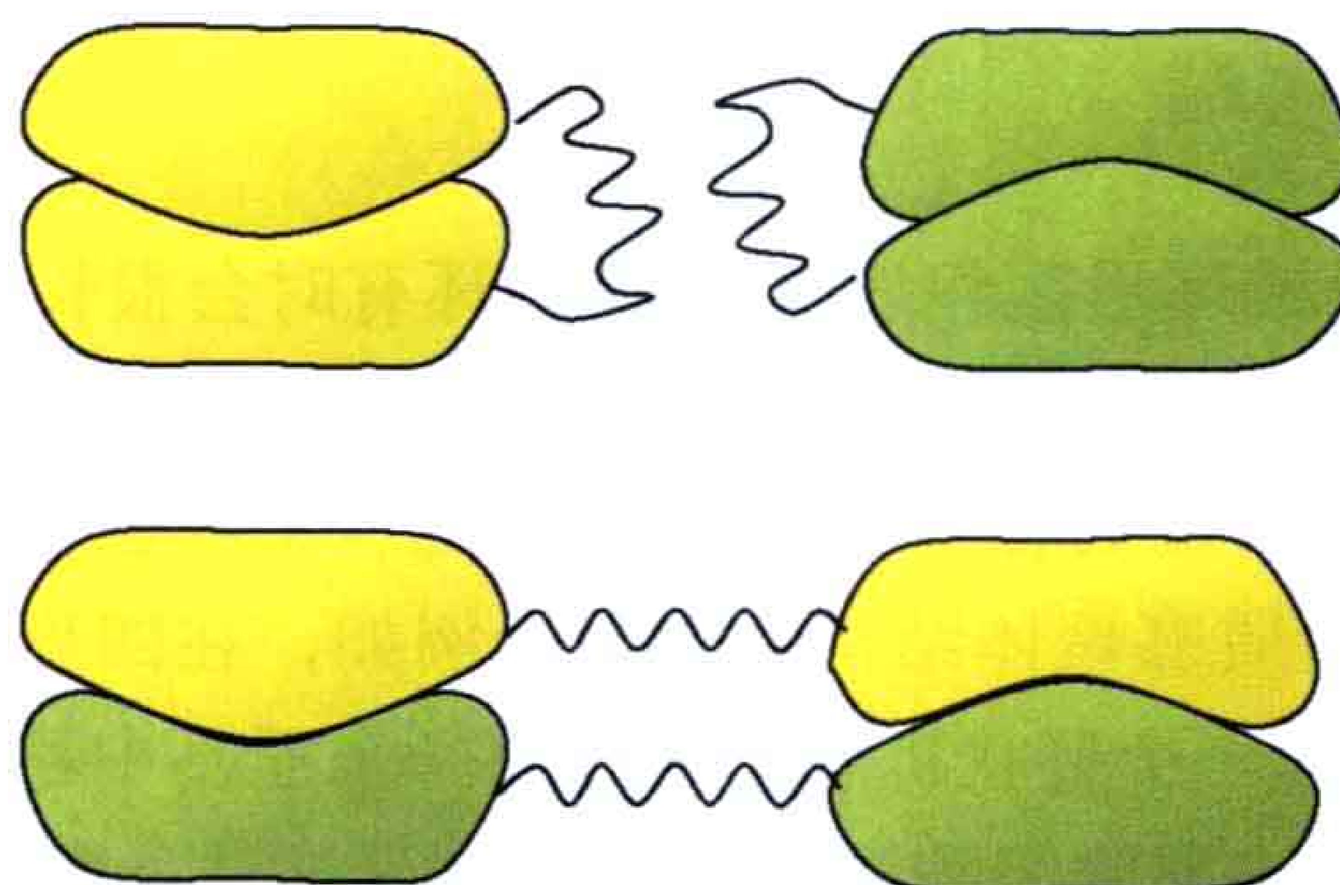


图 2.26　蛋白质的结构域互换。上：正常情况下蛋白质是一个单体，两个结构域在单体间相互作用；下：蛋白质的两个结构域以同样的方式作用于不同的单体形成二聚体。

有时这种方式会延伸成为线性的甚至分支的多聚体。最后的结果是导致单体的聚集。

2.1.6.3　镶嵌型蛋白

一些真核蛋白（也常被称为“镶嵌型蛋白”）由许多模块组成。这些模块也是功能单元，能够从序列模式中认定出来。每个模块在序列中是一个分立的单元并折叠成为独立结构域。例如，组织血浆酶原激活因子由 5 个 4 种不同的模块组成（图 2.27）：一个 FnI（I 型纤连蛋白结构域）模块、一个 EGF（表皮生长因子）模块、两个 Kringle 模块，以及一个主要酶部分，即丝氨酸蛋白酶模块。血浆酶本身也是一个镶嵌型蛋白，由 5 个 Kringle 模块及随后的丝氨酸蛋白酶模块构成。在一些镶嵌型蛋白中，相似模块的个数是很多的（可达几十到上百个——译者注）。其中的一些蛋白质模块，如 EGF 模块和 Fn III（III 型纤连蛋白结构域）模块在大量蛋白质中被发现，并且这些模块似乎都执行着类似的功能（第十一及十三章）。这些模块的大部分并不具有催化功能，而是结合其他分子。



图 2.27　两个典型的多结构域蛋白的结构域组成。上：组织血浆酶原激活因子；下：一个受体型酪氨酸激酶，来源于 Pfam 数据库（http://pfam.sanger.ac.uk），柱状体是相应家族的链接。

镶嵌型蛋白在演化进程中由遗传元素的重组及不同的剪切事件形成。每个结构域可能对应于 DNA 中的单个外显子。基因复制及这些外显子的不同剪切将导致形成由这

些外显子及结构域全新组合的蛋白质分子。

2.1.6.4 环形区及尾端

二级结构元素由转角或环形区相连接。这种环有时会很长，并且会从更紧致的蛋白质中伸出。这些环可能完全缺乏二级结构或者可能形成β条带或螺旋。其他从紧致结构中延伸出来的可能是N端或者C端的尾部。这些尾端也可能有二级结构。环形区及尾端常常起着稳定蛋白质寡聚体结构的作用。例如，在四聚体酶乳酸脱氢酶中，N端的20个氨基酸残基形成一个臂状折叠伸入另外一个亚基中稳定二体-二体相互作用（图2.28）。其相关酶马来酸脱氢酶没有这个N端延伸臂，因而常常仅形成二聚体。

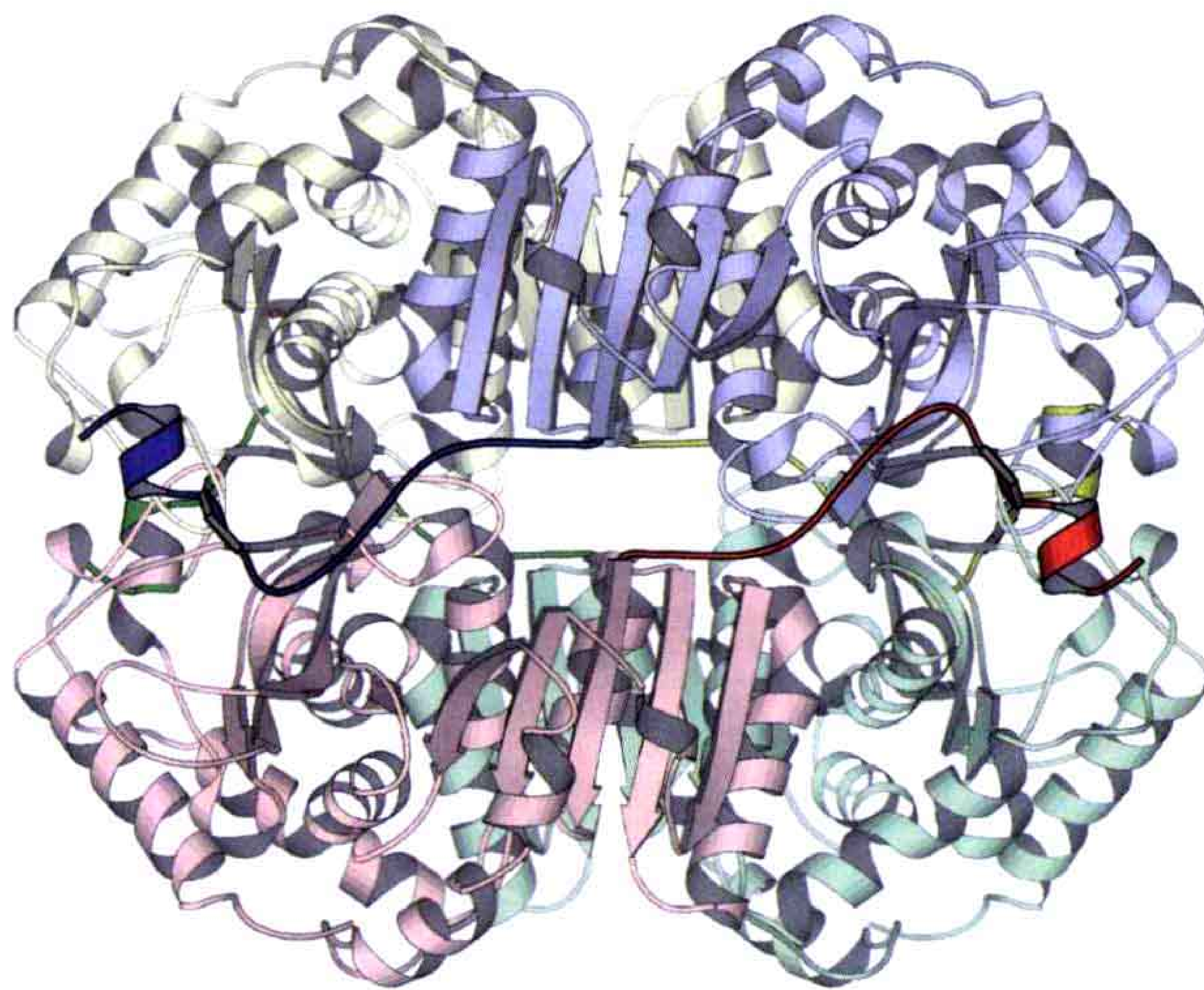


图2.28 四聚体酶乳酸脱氢酶。重点标示N端伸展臂与其他亚基相互作用以保持复合物稳定（PDB:1LDM）。某些物种中相关酶马来酸脱氢酶缺乏这个伸展臂。因而常常形成二聚体。二聚体对应于蓝-绿或者红-黄相互作用。

病毒的包被蛋白（第十五章）、组蛋白（3.2.1节）及核糖体蛋白（第八章）常常具有带正电荷的延伸部分并用其来中和核酸的负电荷。这些延伸元素很可能是这些巨大复合物中的最为古老的蛋白质元素，而且很可能在其他蛋白质中普遍如此。与金属离子相比，这些蛋白质延伸元素可能有变构的优势，几个相连的正电荷可以与核酸中的负电荷结合。球状结构域可能是这些多肽后来演化出的、更高级的特异性及附加功能。

2.1.6.5 天然不折叠的蛋白质

大多数蛋白质具有有序结构，尽管它们可能并不僵硬，但是很多情况下蛋白质中或大或小的部分会有高度的柔性。这个现象的一个解释是柔性区的功能是用来结合其他分子的，而这个结合区域的构象在没有结合伴侣时是不稳定的。结合构象可能是延展的或者是一种拓扑上很难结合的刚体形状。

我们越来越强烈地意识到有一大部分蛋白质天然不折叠或者仅瞬间折叠。一类内禀非结构蛋白（intrinsically unstructured protein, IUP）如胸腺肽 β_4, 结合 G- 肌动蛋白以防止纤维状肌动蛋白的组装（12.1 节）。在其结合形式中，螺旋结合于肌动蛋白单体的反端以防止其聚合。 此蛋白质利用其螺旋为结合元素。在分离出的非折叠蛋白中，那些螺旋仅短暂存在，来降低完全非结构蛋白结合其靶蛋白时的熵损失。

2.1.6.6　纤维状蛋白

纤维状蛋白（第十二章）至少可以三种方式形成。第一类型是由球状蛋白聚集成的线性或者螺旋形式（图 2.29，还可参见 2.2.3.2 节）。

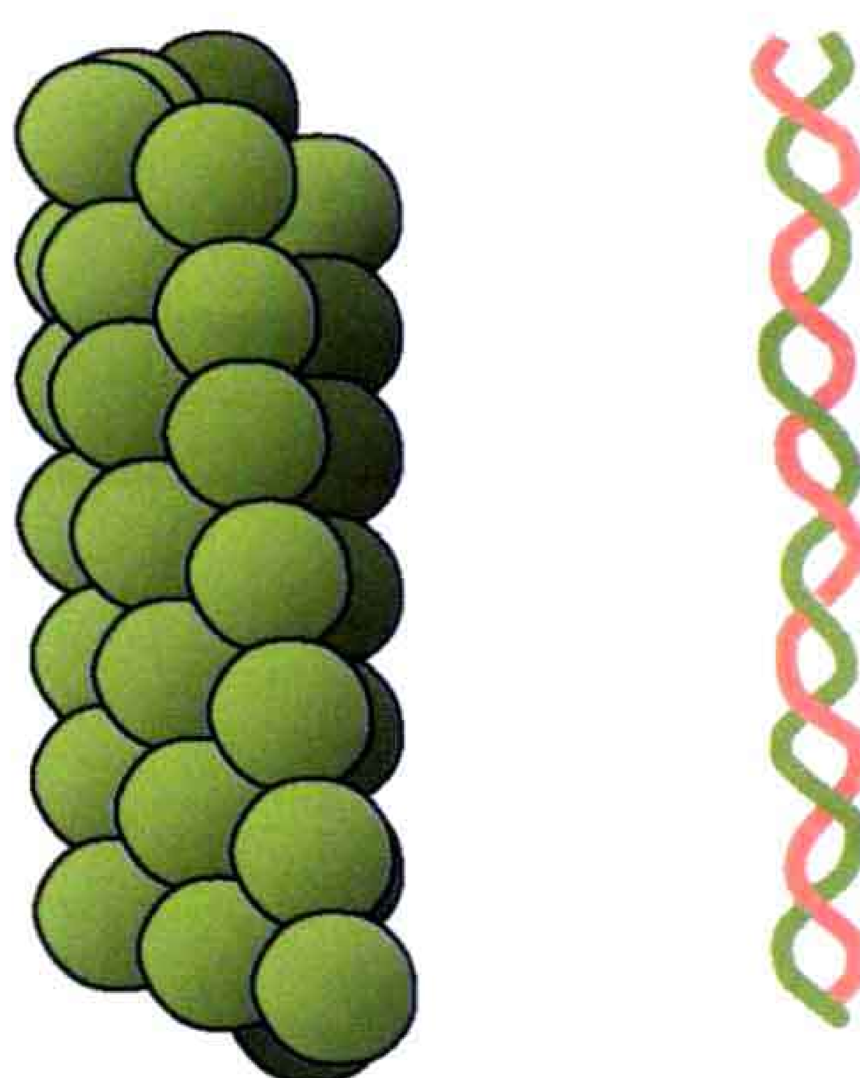


图 2.29　两种纤维状蛋白形成机理。球形亚基聚集呈螺旋状（左），伸展的蛋白质形成盘绕螺旋（coiled-coil）纤维（右）。

第二类纤维状蛋白由非常延伸的并且常伴有特定类型二级结构的多肽链形成（图 2.29）。毛发中的角质蛋白就属于这类蛋白质，它由很长的 α 螺旋构成。 其他这类蛋白质还有肌肉纤维中的肌球蛋白和原肌球蛋白（第十二章），它们由两分子很长的螺旋互相缠绕在一起构成，形成盘绕螺旋结构。

第三类纤维状蛋白如淀粉状沉积纤维。这里大量 β 束产生错误折叠形成很长的反平行 β 片层。 这种情况下，不形成 β 结构的蛋白质部分也有可能形成结构域互换的错误结构（2.2.3.4 节）。

2.1.6.7　膜蛋白

膜蛋白是由好几类蛋白质所组成的（第十章）。很多膜蛋白介导跨膜的信号转导，还有一些涉及离子或者大、小分子细胞膜内外的运输。由于这些生物功能的重要性，膜蛋白常常成为药物靶标。

一类膜蛋白是由一个性质与可溶性蛋白类似的球状部分及固定在膜上的尾部构成，这样的尾部通常是个一次跨膜且与膜垂直的 α 螺旋。其他膜蛋白可以通过好几个与膜

垂直的跨膜 α 螺旋完全被包埋在膜内。一个经典例子是由 7 个跨膜螺旋组成的细菌视紫红质（bacteriorhodopsin）。这类膜蛋白也会有环形区，整个结构域或者亚基延伸到胞质区或者膜外。

α 螺旋并非仅有的一种能够穿过膜的二级结构，也有完全由 β 束构成的整合膜蛋白。这里的典型例子是细菌膜孔蛋白（porin）家族。膜孔蛋白是由很多条反平行 β 束构成的封闭的跨膜圆柱（图 2.15）。形成跨膜圆柱的 β 束数目变化很大。

跨膜螺旋及 β 圆柱的共同基本特征是所有主链极性基团都与其他主链原子形成氢键。这对暴露在膜内疏水区域的极性基团是非常不利的。跨膜螺旋及 β 圆柱蛋白膜外区的非极性侧链会与膜脂相互作用。

还有一类与以上所述不同的膜蛋白只包埋在双层脂膜的一边，如前列腺素 H2 合成酶。这里几条平行于膜的 α 螺旋形成主要的接触区域。

2.2 蛋白质的折叠类型及功能

2.2.1 折叠与功能的关系

2.2.1.1 蛋白质分类系统

测定蛋白质的结构比测定 DNA 序列要付出更多的努力，尽管如此，近年来蛋白质新结构仍然以高增长率产生。随着蛋白质构象（结构）数据库的大幅度增长，需要对蛋白质结构进行有效的分类，即确定蛋白质结构（更确切地说是蛋白质结构域）的分类学。蛋白质氨基酸序列间的巨大差异并没有直接导致折叠类型的千差万别。许多蛋白质尽管没有显著的演化相关性（序列同源性——译者注），但仍然具有相同或者类似的折叠类型。正如对折叠类型的定义一样，对蛋白质折叠类型进行分类的各种方法多少有些主观性。一个较为常用的分类办法是根据其所含二级结构将所有蛋白质分为 5 个主要的组或者级：全 α、全 β、α/β、α+β 及第五组没有多少二级结构从而不适合上述任何类型的蛋白质（附录 B）。α/β 和 α+β 蛋白的区别在于 α/β 蛋白主要由平行 β 束及相连接的螺旋组成（βαβ 模体），而 α+β 蛋白的 β 束及螺旋区域间的连接没有定规。上述每个折叠组中包含许多不同的折叠类型，而每个折叠类型中还有不同的变种。

2.2.1.2 生化功能：趋同与趋异

应该意识到，对于蛋白质结构的折叠类型分类并不必与功能分类相关联。一般来说，某一特定功能并不会局限于特定的折叠类型。丝氨酸蛋白酶及相关酶类很好地说明了这一点，它们是好几种具有类似功能的不同家族蛋白，却有着完全不同的折叠类

型。丝氨酸蛋白酶糜蛋白酶属于真核蛋白水解酶为主的一大类酶。它由两个相似的β桶折叠结构域组成（图 2.30）。糜蛋白酶的活性中心及催化机理都与枯草芽孢杆菌蛋白酶（subtilisin，一个细菌蛋白）很相似，然而却具有完全不同的折叠（α+β 类）。糜蛋白酶和枯草芽孢杆菌蛋白酶没有共同起源。尽管其催化活性中心与催化机理非常相似，三个有重要催化机理的残基（催化三联体，Ser-His-Asp）丝氨酸 - 组氨酸 - 天冬氨酸也相同，但却出现在序列中不同地方。这种情形被称为趋同演化（convergent evolution）：同样的生物学功能在两个不同的蛋白质中独立演化出来。丝氨酸蛋白酶的例子显示蛋白质折叠实际上可被看做用来形成稳定的结构骨架，在这个骨架上活性中心及其他生物功能性质可以被塑造出来。



图 2.30　丝氨酸蛋白酶糜蛋白酶（左）及枯草芽孢杆菌蛋白酶（右）的折叠类型是很不同的，但是这两种蛋白质具有相同功能及酶学机理。丝氨酸 - 组氨酸 - 天冬氨酸（Ser-His-Asp）催化三联体如图所示。

现存的蛋白质折叠类型是演化历史的结果，而不是特定类型的蛋白质折叠对于某一类功能的适应性（第十六章），尽管许多功能特点是与特定折叠类型相联系的。例如，免疫球蛋白，所有免疫球蛋白都有好几个结构域，这些结构域都具有很相似的拓扑结构的β三明治类型。这些结构域不仅仅在我们的免疫系统蛋白中存在，也在许多细胞表面受体中被发现。这种相似性的原因可能是大部分或者所有这类蛋白质都是从同一起源演化出来的。

2.2.1.3　常见折叠类型

除免疫球蛋白折叠外，还有许多其他折叠类型存在于很多看起来似乎不相关的蛋白质中。一些常见折叠类型列于表 2.5。其中一些折叠类型的相关蛋白结构图示于图 2.31。

很多酶具有 TIM 桶折叠，见 2.1.5.2 节（名字来源于磷酸甘油醛异构酶，triose phosphate isomerase）。这个折叠由 8 条平行β束形成一个圆筒，每条β束再连接一个螺旋，8 个螺旋组成蛋白质的一个外层，即 $(\beta\alpha)_8$。许多功能不相关且没有显著序列同源性的酶都具有这个折叠，似乎这些蛋白质在演化过程中独立地发展出这类折叠，并

且这些酶的活性中心总是位于桶的同一端。

表 2.5　一些存在于不同蛋白质家族中的常见折叠类型

名称	折叠类型	举例
α/β 双缠绕	主要为平行片层，螺旋分列两边	Ras 折叠、枯草芽孢杆菌蛋白酶、腺苷酸激酶
TIM 桶	8 个 β 束封闭圆柱，间以螺旋相连	磷酸甘油醛异构酶、羟乙酸盐氧化酶、醛缩酶
劈裂 α/β 三明治	反平行片层，螺旋位于一侧	4Fe-4S 铁氧还蛋白、酰基磷酸酶
免疫球蛋白	β 三明治	免疫球蛋白、受体结构域、超氧化物歧化酶
α 上下折叠		血红蛋白，TMV 病毒包被蛋白
球蛋白	两层非平行螺旋	血红蛋白，藻青蛋白
瑞士果冻卷	β 三明治	肿瘤坏死因子，病毒包被蛋白，伴刀豆球蛋白
三叶形	三片层形成的圆柱	白细胞介素，蓖麻毒蛋白
泛素 α/β 圈	小片层，螺旋位于一侧	泛素、4Fe-4S 铁氧还蛋白

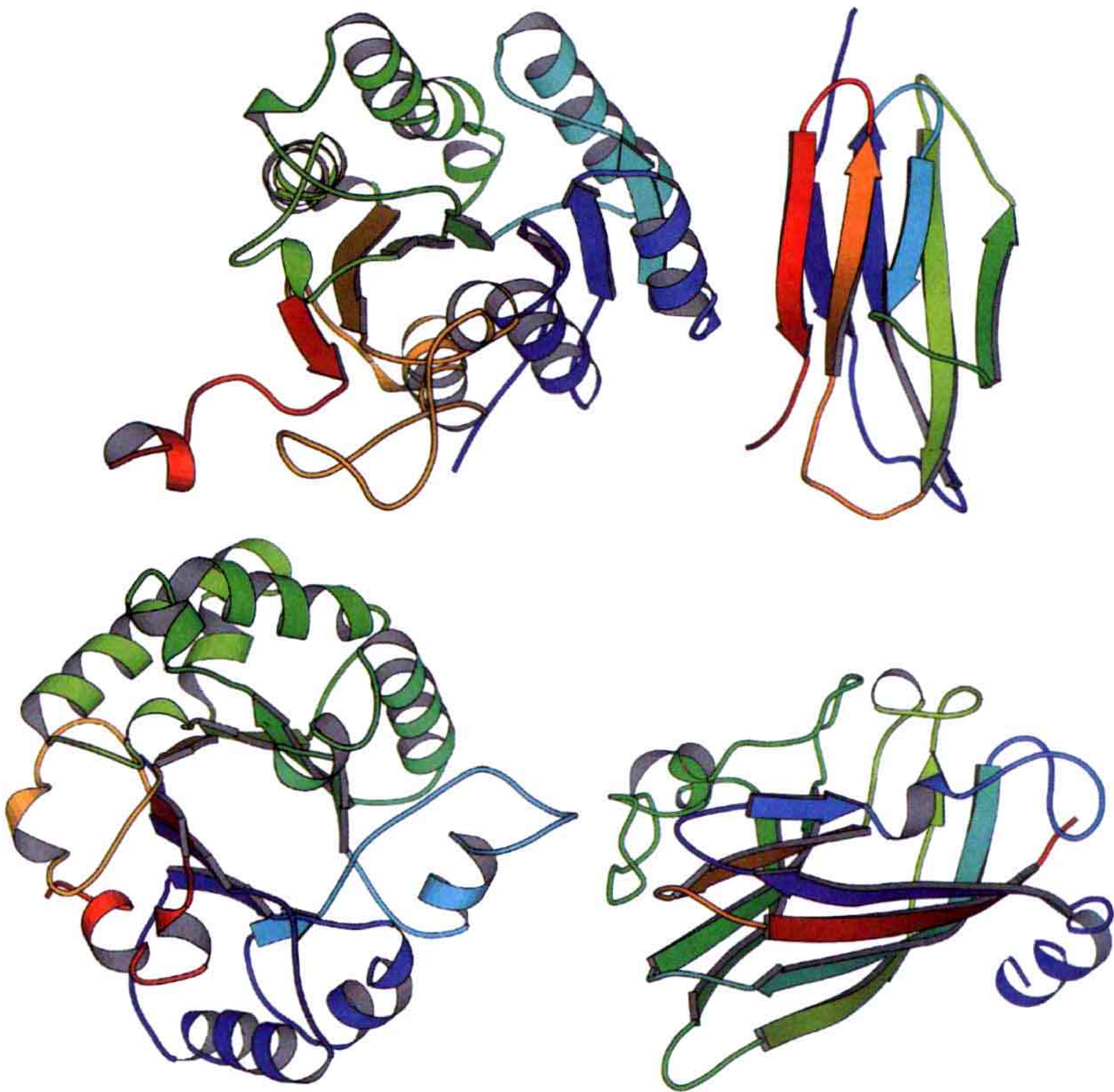


图 2.31　几种蛋白质的共同折叠类型。上左：α/β 结构域的罗斯曼折叠 (3α, 20β- 羟基类固醇脱氢酶 , PDB: 1HDC）；上右：免疫球蛋白不变结构域 （PDB: 1AQK）；下左：TIM 桶折叠 (磷酸甘油醛异构酶 ,PDB: 1YPI)； 下右：果冻卷折叠 (卫星烟草坏死病毒 STMV 包被蛋白 , PDB: 2BUK)。

另一种常见折叠类型是“瑞士果冻卷”（jellyroll）折叠，常见于很多病毒的包被蛋白及其他一些功能完全不相关的蛋白质，如肿瘤坏死因子。

很多蛋白质具有简单拓扑结构的相似折叠。这些功能不相关蛋白中的相似折叠是否为趋异演化而来还不得而知。这样的简单拓扑结构或许在演化中得到偏爱，因为这样的折叠可能具有高效率及精确的折叠过程。

另外，还有好几种折叠类型存在于具有特定功能的某一家族蛋白中。有两个例子：碳酸酐酶与小 RNA 噬菌体的包被蛋白。

随着越来越多的蛋白质结构被测定，观察到的折叠数目在不断增加。来自完整基因组的努力测序而产生出的海量数据使我们能够估计可能存在的蛋白折叠个数。这可以通过把所有序列分类成为不同的同源蛋白组来实现。由于蛋白质构象比序列更加保守，这些组中的每一个蛋白质都具有相同的折叠类型（假设折叠指认是正确的）。很多具有同样折叠类型的蛋白质可能演化相关，但序列发散到没有显著的同源性存留下来，然而其他一些蛋白质可能是真正的非同源并且独立演化出同一折叠类型。从新序列属于已知同源组的比例，我们可外推估计出存在于活细胞中具有显著序列相似性的同源蛋白组的数量。从不同序列组相关联的折叠数目，我们也可估计出这些组中有多少是独特折叠类型。目前的估计是最多只有几千种不同的折叠类型，其中大约 1000 种已知。

大量的但有限的已知折叠数目并不是所有多肽链可能形成的折叠数目（这有时被称为折叠空间）。可能的氨基酸序列总数是非常巨大的。对于一个仅有 100 个氨基酸的小蛋白来说，具有 20^{100}（约 10^{130}）可能的序列。由于相对僵硬的主链以及形成堆积好的疏水核心要求很高强烈地限制着蛋白质的折叠，也许仅有极小一部分理论上可能的氨基酸序列会折叠成为稳定的球状结构。也很有可能许多稳定的蛋白质折叠类型并不存在于自然界，因为现存的折叠是生物演化过程的结果，这个过程并没有搜索或者保存所有可能蛋白质序列的组合。

2.2.2　蛋白质稳定性及动力学

2.2.2.1　折叠过程

蛋白质三级结构域构象是由其氨基酸序列决定的，从而也是由相应基因的碱基序列决定的。构象与蛋白质氨基酸序列的直接关系最早是由安芬森（Anfinsen）实验证明的，具有功能的葡萄球菌核酸酶可以从变性的蛋白质重折叠得到。我们现在对氨基酸侧链怎样形成正确构象的过程，即折叠过程比那时要知道的多得多。

即使对于一个很小的蛋白质来说，所有理论上可行的多肽链构象的数目也是巨大的。尽管如此，蛋白质的折叠发生得很快，这意味着多肽链不可能历经所有的构象来找到最稳定的。这一现象的一个可能解释是折叠过程是沿着一条或者有限的几条路径到达最终折叠状态。蛋白质折叠在生长的多肽链上就已经开始，即“伴翻译折叠”。因此，不能排除折叠过程本身及其动力学对于最终构象是很重要的，也就是说不仅天

然构象而且折叠过程本身都是被氨基酸序列编码的。

对大部分蛋白质来说，天然构象似乎是最稳定的构象状态，尽管折叠及非折叠状态之间的能量差别一般很小。氢键的形成（附录 A.3.1）是蛋白质折叠的重要方面。氢键形成为蛋白质折叠提供关键的能量贡献，并且氢键具有的方向性很大程度上决定了蛋白质的折叠。只要损失几个氢键就足以破坏蛋白质的折叠。疏水作用对于蛋白质折叠也是非常重要的。蛋白质的疏水部分通常被包埋在内部，而带电荷与极性的基团则暴露在表面上（附录 A.4）。

蛋白质天然折叠可能并不代表最低的能量构象，因为折叠过程可能不允许任何路径达到这个状态。在这样的情况下，天然状态是折叠过程允许到达的最低能量构象。折叠好的蛋白质与非折叠蛋白状态之间的微小能量差也是蛋白质行使生物功能时的重要性质。蛋白质功能经常取决于它的某种柔性。更进一步来说，对各种生命过程的控制也需要蛋白质具有一定的寿命并且可以被降解。非常稳定的蛋白质构象与这样的需要不相容。

折叠路径似乎在不同蛋白质之间差别很大。一些蛋白质折叠似乎不需要任何中间过程（中间体）。其他一些蛋白质折叠涉及中间状态，有时称为“溶球状态”，其中仅某些天然相互作用被建立起来，整个完整紧致的折叠状态还没有达到。例如，蛋白质二级结构可能已经形成，但是在中间体中这些二级结构优化的堆积还没有实现。

细胞中还有很多被称为“分子伴侣”（chaperon）的蛋白质，它们帮助某些蛋白质的折叠（9.1 节），分子伴侣通常对于特定折叠没有特异性。伴侣蛋白（chaperonin）的功能，如细菌的 GroEL/ES 复合物及热激蛋白与非天然折叠的蛋白质结合从而防止其聚集，以利于其重新折叠。因此，这类蛋白质的功能是提高蛋白质折叠效率而不是指导特定构象的折叠过程。

有些蛋白质只能以前体蛋白的形式正确折叠，折叠完成以后多肽链的一段会被去除。一个著名例子是小蛋白激素胰岛素。去除小肽的功能可能是除去折叠途径中的能障。

2.2.2.2 蛋白质堆积及蛋白质中的空洞或空腔

对于蛋白质的稳定性来说，二级结构的堆积最为重要。使蛋白质折叠的各种相互作用力与形成二级结构的相互作用没有什么不同，主要是共价键、氢键、静电相互作用及范德瓦耳斯力。另外，疏水作用也贡献巨大。与非折叠多肽链结合的水分子在折叠过程中大多会释放出来，使熵得到补偿，稳定蛋白质。有些水分子也会存留于蛋白质结构中（详见 2.2.4.2 节）。然而，也会有一些小的空洞或者堆积缺陷依然存留于结构中，许多这类空洞太小不会结合任何东西。有些情况下，空洞会较大，为 25 ~ 150Å^3。惰性气体如氙气或者氪气能够结合在这些较大的空洞中。

由此类空洞而导致的蛋白质稳定性的损失，在细菌噬菌体 T4 的溶菌酶中进行过详细的研究。对一系列的疏水核心突变进行详查，将较大的疏水侧链换成较小的，如丙氨酸。尽管这样产生出的空洞周围的残基移到了空洞中一些，但这些突变多数不会改

变蛋白质的整体结构。没有观察到水分子结合于空洞中。稳定性的损失为 100 ~ 140 J/mol/$Å^3$ 每空洞体积，或者 80 J/mol/$Å^3$（按照空洞表面积计算）。

蛋白质结构是其功能的主要决定因素，然而其动力学也起着关键作用。蛋白质需要采取不同的构象，因而部分蛋白质必须能够移动到新的位置，在一个僵硬的分子中这是很难做到的。但是，如果蛋白质中存在空间从而允许不同的部分采纳不同的构象，运动则较少受到限制。因此，蛋白质中的空洞具有重大的功能意义。这点在肌红蛋白中应用不同的方法包括一系列定点突变进行了广泛深入的研究。利用蛋白质中的空腔，配体分子可以传输进入或者离开血红素基团，并且 CO 配体在这些空腔中的结合在光解离后可由光谱技术及时间分辨的晶体学方法（劳厄晶体学）追踪。对于影响空腔的突变蛋白的研究显示，配体传输的速率取决于空腔核心的堆积。这个例子显示蛋白质的氨基酸序列不仅仅编码了蛋白质的折叠，而且也影响着其动力学性质。

2.2.2.3　变性——热稳定性

蛋白质的稳定性很弱（处于边缘）。折叠状态的稳定性可被描述为两个符号相反的大量热力学焓及熵之间差别的平衡。因此，蛋白质很容易被各种方法变性，如加入变性化学试剂（尿素、盐酸胍等）、变化 pH，或者提高温度。另外，二硫键 S—S 或者金属离子的偶联效应则能够稳定蛋白质。

比较来自不同物种的具有同样功能但不同氨基酸序列的蛋白质，显示出这些蛋白质可能在上述变性实验中具有不同的稳定性，这可能是这些物种不同生活环境的反映。因此，“嗜热”（thermophilic）物种的酶普遍对于热变性更加稳定。另外，这些酶在低温下不如相应“嗜温”（mesophilic）物种的酶活性高，“嗜温”物种的酶的活性又不如“嗜冷”（psychrophilic）物种的酶高。工业过程往往在高温下进行，理解好热变性是工业利用酶的重要前提。一个中心问题是怎样才能够改进酶的热稳定性。突变分析表明蛋白质中的空腔会破坏稳定性，因此，将这些空腔用较大的侧链充满会增强稳定性。另一种稳定因素是增加氢键及电荷 - 电荷相互作用的数目。引入二硫键也能够增加稳定性。

蛋白质折叠有时被描述为“拼图游戏”（jigsaw puzzle）。这一描述在解释蛋白质对于表面氨基酸有很高的替换容忍度上遇到一定程度的挑战。尽管疏水核心的氨基酸替换也可以被容忍，但这常常导致稳定性降低。

布莱恩 • 马修斯（Brian Matthews）及其同事在 T4 溶菌酶中对蛋白质的稳定性进行了非常深入细致的研究，对很多突变体的变性温度、酶活性及结构进行了系统分析。如上所述（2.2.2.2 节），T4 溶菌酶的堆积缺陷相应于稳定性降低被详细探究过。对活性中心的催化残基或者底物结合相关残基进行系统突变可以研究蛋白质稳定性和功能之间的可能冲突。结果表明，这些残基的突变可以使蛋白质以降低活性的代价来提高稳定性。类似地，也进行了重新设计疏水核心的尝试。许多疏水残基如亮氨酸、异亮氨酸、缬氨酸或者苯丙氨酸被替换为甲硫氨酸，其占据大约相同的体积，但是具有更高的柔性可以采取合适的构象。研究了多至含有 10 个甲硫氨酸替换的突变蛋白，突变可以被容忍，但是蛋白质的稳定性降低了。

2.2.3 蛋白质 - 蛋白质相互作用

生物大分子的功能取决于它们的结构及其相互作用。不仅仅是较小的分子及离子，大分子间的相互作用也在很大程度上也决定了其功能。蛋白质能够形成大的寡聚体，或者由一种构筑单元组成（同源寡聚体，homo-oligomer），或者由多种单元组成（异源寡聚体，hetero-oligomer）。要形成稳定的相互作用，大分子的相互作用表面（以溶剂可接触表面积的减少来测量——译者注）需要足够大。稳定复合物的平均相互作用表面大小约为 1600 $Å^2$。

2.2.3.1 瞬态及长时相互作用

寡聚体蛋白通常是具有很长寿命的稳定分子聚集体。一个寡聚体酶的功能实现可能并不需要亚基间的相互作用而是由于寡聚体具有更高的稳定性。另外，某些寡聚体酶亚基间具有变构相互作用，其可调控蛋白质功能（第五章）。稳定寡聚体，如病毒外壳或者纤维状肌动蛋白，具有动态特性，能够由单体多聚化及去多聚化成为单体来实现特定功能。

许多蛋白质的相互作用是短时的，酶即属于此类蛋白质（第五章）。它们结合各自的底物，有时底物是大分子，实现酶的催化活性然后解离。在此过程中底物或者酶会改变构象，导致分子间亲和性（亲和系数）降低。很多没有酶活性的蛋白质也会采取类似的大分子间暂时相互作用。例如，信号转导通路中的接头蛋白会把相关蛋白结合在一起使某类酶活性施加在合适的底物上（第十一章）。其他例子如 ATP 酶和 GTP 酶通常作用于各自的核苷酸交换因子及相关的 ATP 酶或 GTP 酶活化蛋白（第五章）。第三个例子是帮助蛋白质合成不同步骤中的翻译因子在短时间内结合及解离核糖体（第八章）。所有这些蛋白质都具有瞬时相互作用并且对于控制其他蛋白质的活性非常关键。

2.2.3.2 寡聚体蛋白，对称及对称破缺

蛋白质常会倾向于聚集。这有时会导致具有很好功能的大复合体的形成；然而有时也会形成纤维样聚集，会对机体造成伤害。一个很重要的蛋白质聚集领域是用于结构解析的蛋白质晶体学。

蛋白寡聚化的各种变化排列似乎具有无限可能性。相同蛋白质分子相互作用时通常利用不同蛋白质亚基（原聚体）上的相同表面。这常会导致一个固定分子数目封闭排列或者许多分子间具有相同相互作用的开放排列。这样的排列通常会有某种类型的对称性（图 2.32）。

依据相互作用表面的空间关系，蛋白质寡聚体可以形成线性（较少见）、螺旋型或者封闭型物体。很多研究得很清楚的结构具有螺旋对称性。一些螺旋聚体的例子列于表 2.6。

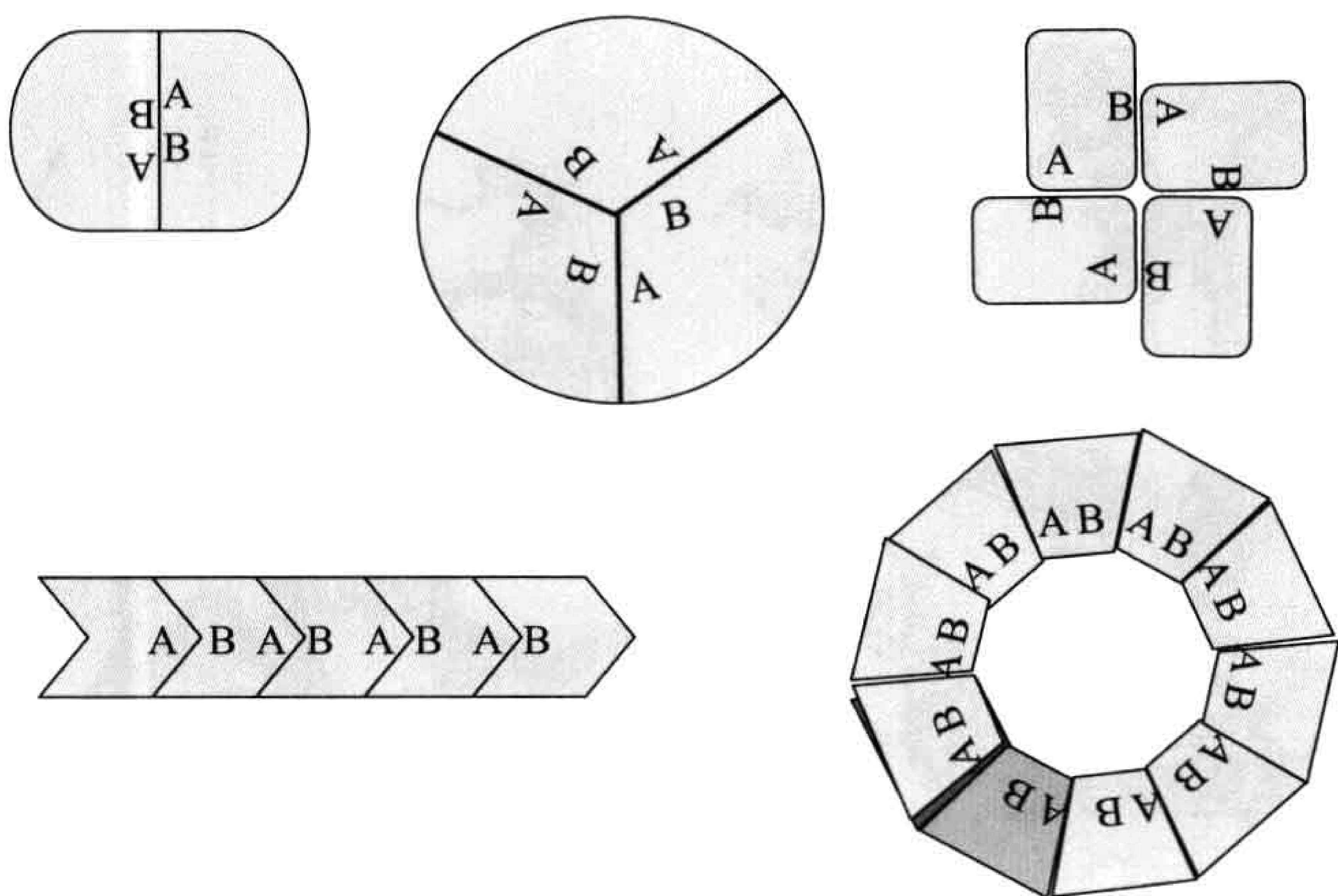


图 2.32 蛋白质寡聚体的一些例子。取决于亚基（A 和 B）相互作用表面的相对取向，蛋白质寡聚体可以形成二聚体、三聚体、四聚体、线性排列或者螺旋排列寡聚体。

表 2.6 螺旋聚集蛋白的一些例子

蛋白质	单体螺旋上升距离 /Å	单体间旋转角 /°	说明
肌动蛋白	27	167	–
IV 型纤毛	10.5	100	–
纤维状噬菌体，fd	32	72 及 180	具有相反取向的五聚环
烟草花叶病毒 (TMV)	1.4	22	–

封闭型结构可以具有一个或几个对称轴。一个具有二重对称轴的简单二聚体是形式最简单的蛋白质寡聚体。有的生物系统具有多达十七重旋转对称（图 2.33 和表 2.7）。

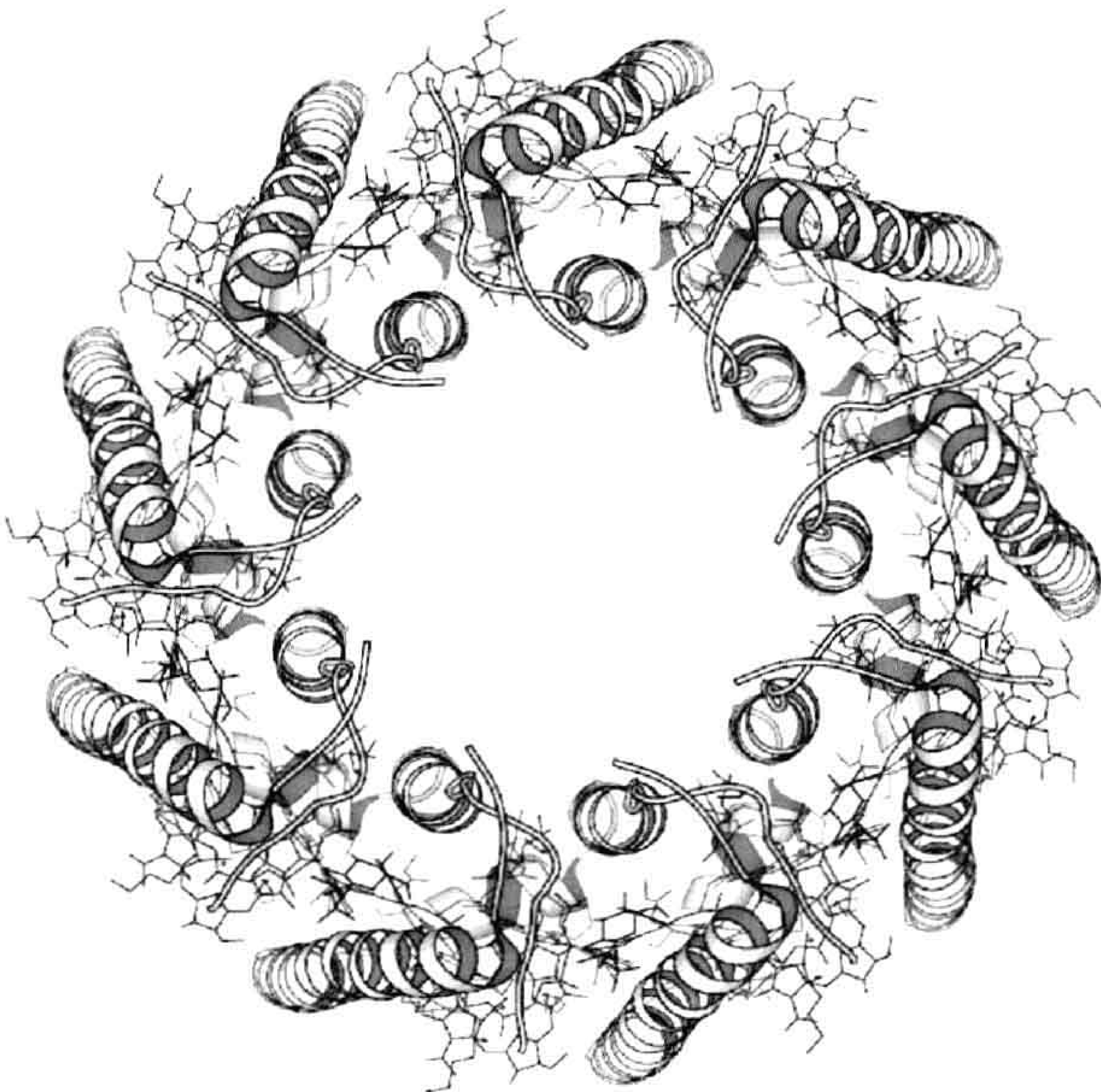


图 2.33 *Rhodopseudomonas acidophila* 的光俘获复合物。具有九重对称性的一个分子 (PDB: 1IJD)。

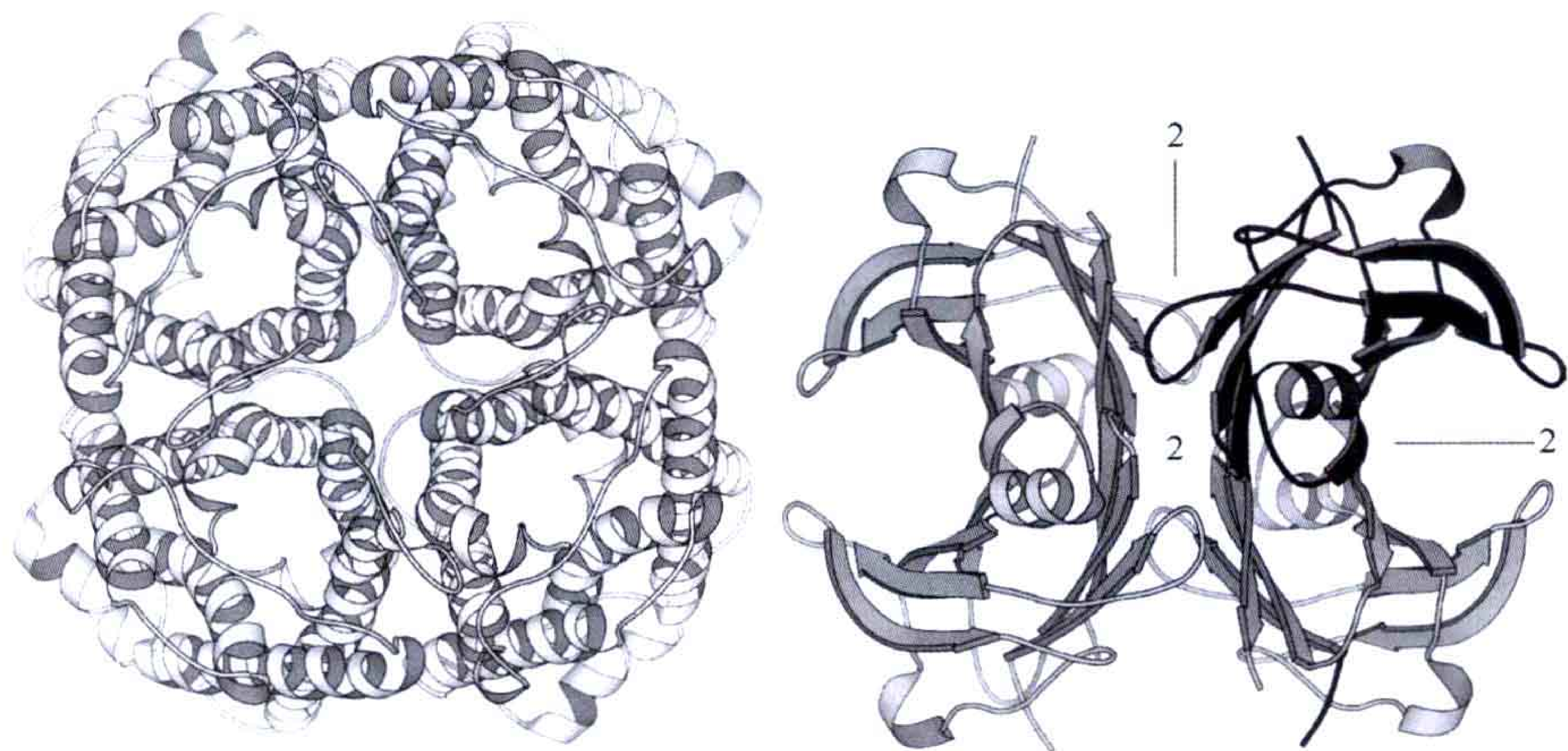


图 2.34 两种四聚体的不同堆积。左：水通道蛋白(四重旋转对称,PDB:1J4N)；右：甲状腺素(222对称,PDB:1OO2)。

表 2.7 周期对称聚集的蛋白质

旋转对称	举 例	说 明
2	乙醇脱氢酶	–
3	膜孔蛋白，流感病毒凝血蛋白	–
4	流感病毒唾液酸苷酶，水通道蛋白	–
5	血清淀粉样 P 组分	–
6	RepA 解旋酶	–
7	GroEL,GroES	七次及二次对称
8	光俘获蛋白复合物 2（LHC2）	–
9	光俘获蛋白复合物 2（LHC2）	–
11	Trp RNA 结合弱化蛋白（TRAP）	–
16	光俘获蛋白复合物 1（LHC1）	–
17	TMV 包被蛋白环	双环

最为复杂的寡聚体对称性包括不同方向上的多重对称轴。最简单的这类排列为具有四聚体对称性的三个互相垂直的二重轴(222 对称性，图 2.34)。已经发现三重、四重、六重和七重对称性与相互垂直的二重轴组合的例子。最复杂的复合物常常是不同形式的立方对称性，如正二十面体病毒具有 532 对称性（表 2.8）。立方对称性是由 4 个沿着立方体对角线方向的三重轴而产生的。

表 2.8　具有多重对称轴的蛋白质复合物

对称性类型	举　例	说　明
222	乳糖脱氢酶	4 个亚基
32	天冬氨酸甲氨酰转移酶	两种 12 个亚基
422	蚯蚓血红蛋白，葡糖酸氧化酶	8 个亚基
622	谷氨酰胺合成酶	12 个亚基
23	Protechuate 双加氧酶	四面体对称，12 个亚基
432	铁蛋白	八面体对称，24 个亚基
532	正二十面体病毒	二十面体对称，60 个或者 60 的倍数个亚基

2.2.3.3　多聚化 - 去多聚化

一些功能重要的蛋白质根据生理需要经历多聚化－去多聚化反应，如肌肉中的蛋白质——肌动蛋白（12.1 节）。这个蛋白质有两种形式，即 G- 肌动蛋白和 F- 肌动蛋白。F- 肌动蛋白为聚集形式或称为丝状形式，这取决于与其他大量蛋白质的相互作用，它可以形成螺旋状多聚体或者去多聚化成为单体。 肌动蛋白在肌肉细胞中形成细丝。 另一种蛋白质，即微管蛋白，可以被其他蛋白质调控实现多聚化 - 去多聚化，从而在细胞中形成微管。

很多病毒的外壳包被蛋白也可在病毒感染宿主时去多聚化，而在它们准备好离开宿主细胞感染其他新细胞时多聚化形成新的病毒颗粒。

2.2.3.4　淀粉样聚集

要执行其功能，蛋白质必须适当地折叠。 分子伴侣系统可以避免蛋白质的错误折叠（9.1 节）。 某些蛋白质的特定突变体倾向于折叠成为导致该蛋白质聚集成不可溶的形式，这些不可溶纤维称为淀粉样聚集（amyloid）。这种沉淀可能导致很多疾病状态 (表 2.9)，这些疾病状态的另外一个名字是蛋白质错误折叠紊乱（protein misfolding disorder，PMD）。 这类纤维沉淀一般由一种蛋白质多聚化构成。能够形成淀粉样聚集的蛋白质一般没有相关性。结构研究表明，淀粉样纤维都有共同的特殊纤维衍射图形，称为交叉 β 衍射花样（图 2.35）。赤道面上相应于大约 10Å 的衍射峰是由于 β 片层的堆积形成的。 根据氨基酸的组成不同，片层间的堆积排列成不同的间距。 在与此相互垂直的方向上（沿着纤维束轴）有一个相应 β 束间距为 4.7 ～ 4.8Å 的衍射峰。图 2.36 显示一个会产生出交叉 β 衍射花样的平行扭转 β 片层结构。

表 2.9　蛋白质错误折叠疾病

疾　病	蛋白质
阿尔茨海默病（AD）	Aβ 多肽
CJD （Creutzfeldt-Jakob's disease）	Prion （普里昂蛋白）
帕金森病 (Parkinson disease)	α-Synuklein
亨廷顿病 (Huntington disease)	Huntingtin

续表

疾　　病	蛋白质
II 型糖尿病	胰岛淀粉样多肽 (IAPP)
家族性淀粉样多发性神经病	甲状腺素
老年全身性淀粉样变性病	甲状腺素
遗传性系统性淀粉样变性病	溶菌酶
芬兰型家族性淀粉样变性病	凝溶胶蛋白
轻链淀粉样变性病	免疫球蛋白 VL 结构域
英国家族老年痴呆症	ABri
主动脉中的淀粉样蛋白变性	Medin
继发性系统性淀粉样变性病	Serum Amyloid A（血清淀粉样蛋白 A）
spinocerebrellar 共济失调症	Ataxins 1, 3, 7
血液透析相关淀粉样变性病	β-2 microglobulin （β-2 微球蛋白）

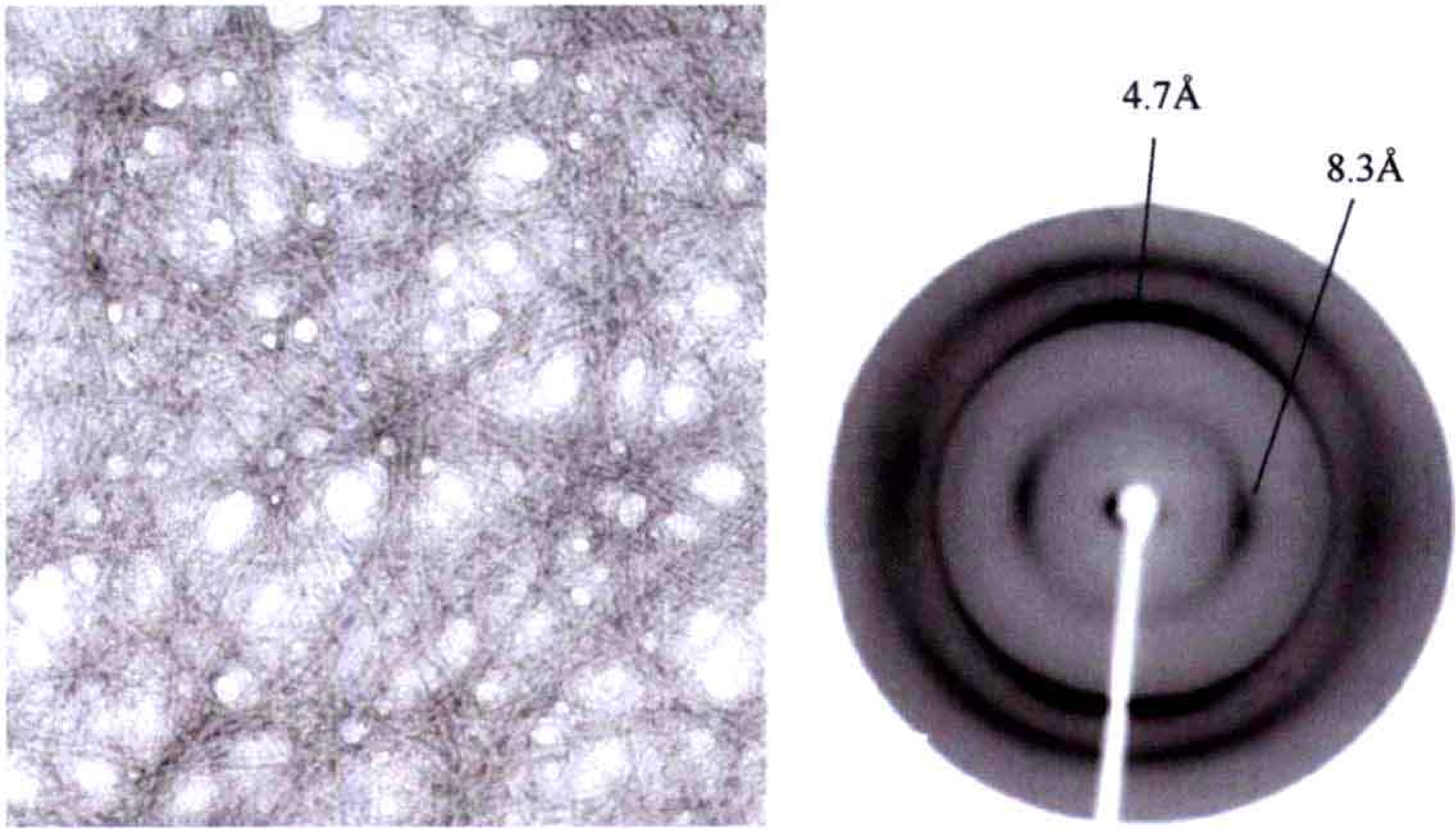


图 2.35　左：IAPP（残基 1~37）的淀粉样聚集纤维（由 G. Westermark 慷慨提供）。右：淀粉样聚集纤维的交叉 β 衍射花样的一个例子。4.7 Å 衍射的距离对应于 β 片层中束与束的间距。8.3 Å 的衍射可变，取决于堆积片层之间的距离[得到许可重印自 Bader R, Seeliger MA, Kelly SE, Ilag LL, Meersman F, Limones A, Luisi BF, Dobson CM, Itzhaki LS. (2006) Folding and fibril formation of the cell cycle protein Cks1. *J Biol Chem* **281:** 18816-18824. Copyright (2006) Elsevier.]。

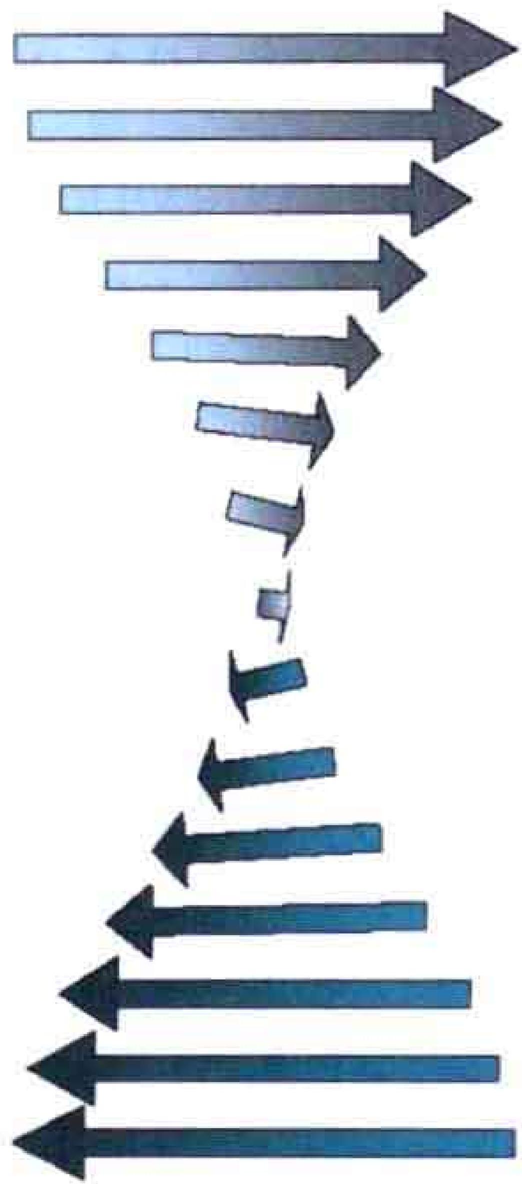


图 2.36　聚集成为平行 β 结构的多肽产生典型的交叉 β 衍射花样。

如果一个具有 β 结构的蛋白质，其片层的一侧没有被侧链或者其他结构所阻挡，其 β 片层就可以自由延伸。 运甲状腺素蛋白（transthyretin）就是一个这样的例子，它是能够导致PMD的很多蛋白质中的一种。有时淀粉样聚集不仅仅涉及β 束间相互作用，如果这个 β 束是某个蛋白质两个结构域间连接的部分，蛋白质结构域也可以失控结构域互换（runaway domain swapping）的方式成为纤维聚集（图 2.37）。

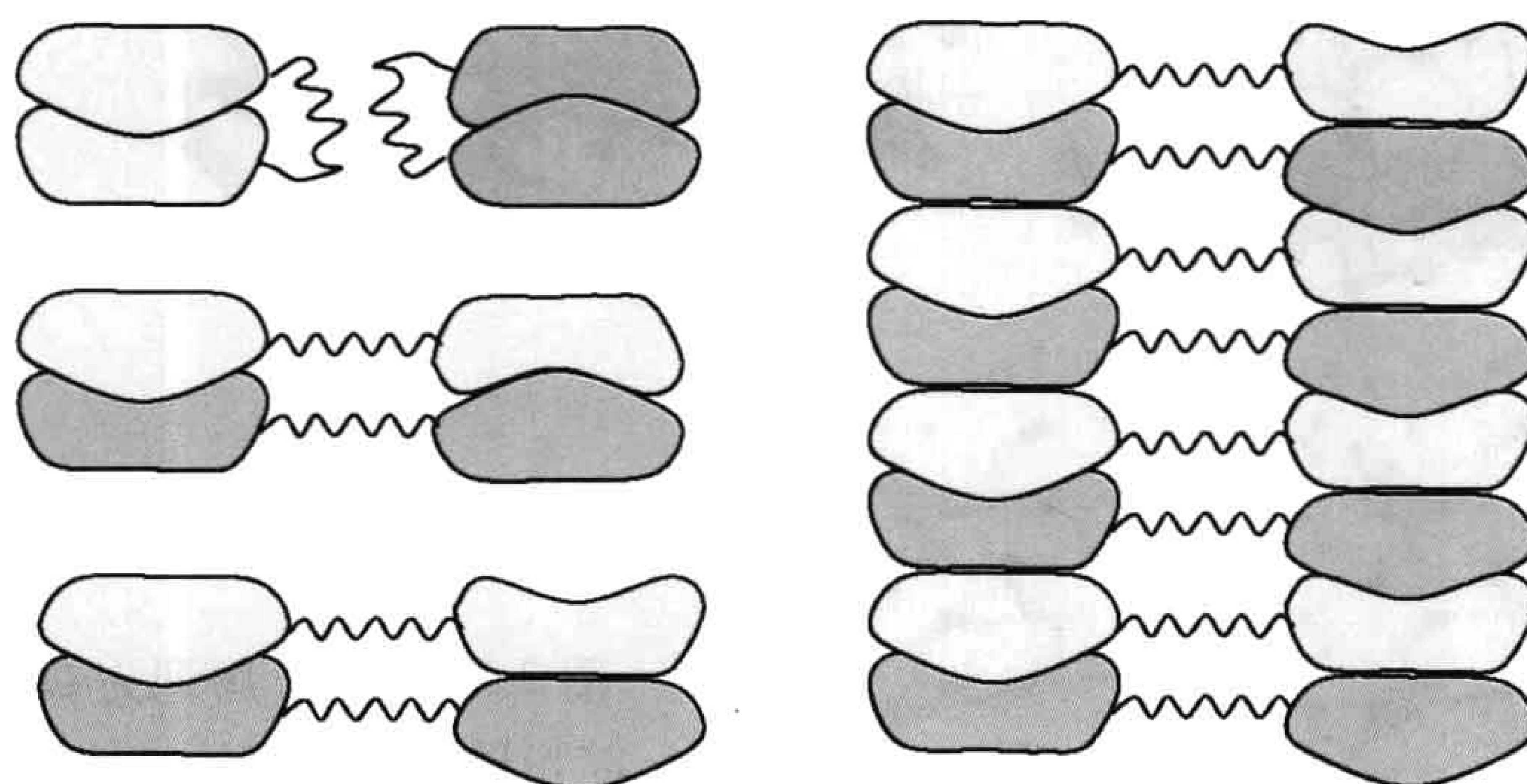


图 2.37　失控结构域互换（runaway domain swapping）现象的一种描述。上左图中显示了两个分立的单体蛋白。 它们可以两种不同的方式形成结构域互换的二聚体（下左图）。与此同时，两个结构域间的连接部分形成一个 β 片层。 这样的两种方式交互形成二聚体，进而导致蛋白质通过 β 片层及结构域互换相互作用聚集成为纤维。

普里昂（感染性蛋白质，prion）疾病如 TSE（transmissible spongiform encephalopathy）或者疯牛病（mad cow disease）都与 PMD 有关。 蛋白质 PrP^{C}，特别是其突变形式会聚集。 导致疾病的聚集形式是 PrP^{Sc}，它会诱导可溶性的 PrP^{C} 改变构象成为 PrP^{Sc} 并因此聚集。这些普里昂蛋白与淀粉样纤维形成蛋白相似，它们形成广泛的 β 结构并因此聚集。普里昂具有传染性，因为一个个体中的聚集物质也可以在另一个健康个体体内诱导聚集。 然而，淀粉样聚集是否普遍性地具有传染性还是一个未解决的问题。

某些多肽已被确认为形成淀粉样纤维的罪魁祸首。 其中的一些结构研究给出了交叉 β 结构形成的细节。 在酵母普里昂蛋白 Sup53 中，七肽 GNNQQNY 及六肽 NNQQNY 引起淀粉样聚集。 这些多肽的晶体结构具有交叉 β 的排列，多肽形成与纤维轴垂直的平行 β 束（图 2.38）。 β 束的间距为 4.9Å。多肽氨基酸残基完美地对正排列，不仅主链而且侧链也互相形成氢键。 如此形成的 β 片层间以两种不同的界面相互作用。 其中一种为干界面 （β 片层之间没有水分子），层间距为 8.5Å。另一种湿界面间含有许多水分子，间距为 15Å。干界面通常排列得很紧致，层间具有广泛的互补性，这可能是淀粉样纤维的一个共性。

淀粉样纤维的结构与昆虫和蜘蛛生产的丝纤维类似，其 β 束顺着纤维轴延展且 β 片层的氢键与纤维垂直。 丝纤维具有大量重复序列 GAGAGS 或者 GAGAGY。有模型建议丝纤维中由于多肽链 β 转角的存在形成了反平行 β 结构（图 2.39）。β 片层的相互作用一方面通过甘氨酸残基进行，而另一方面由丙氨酸、丝氨酸或者酪氨酸的侧链介导，产生片层间不同的距离变化。很显然，丝纤维与淀粉样纤维具有结构上的相关性。

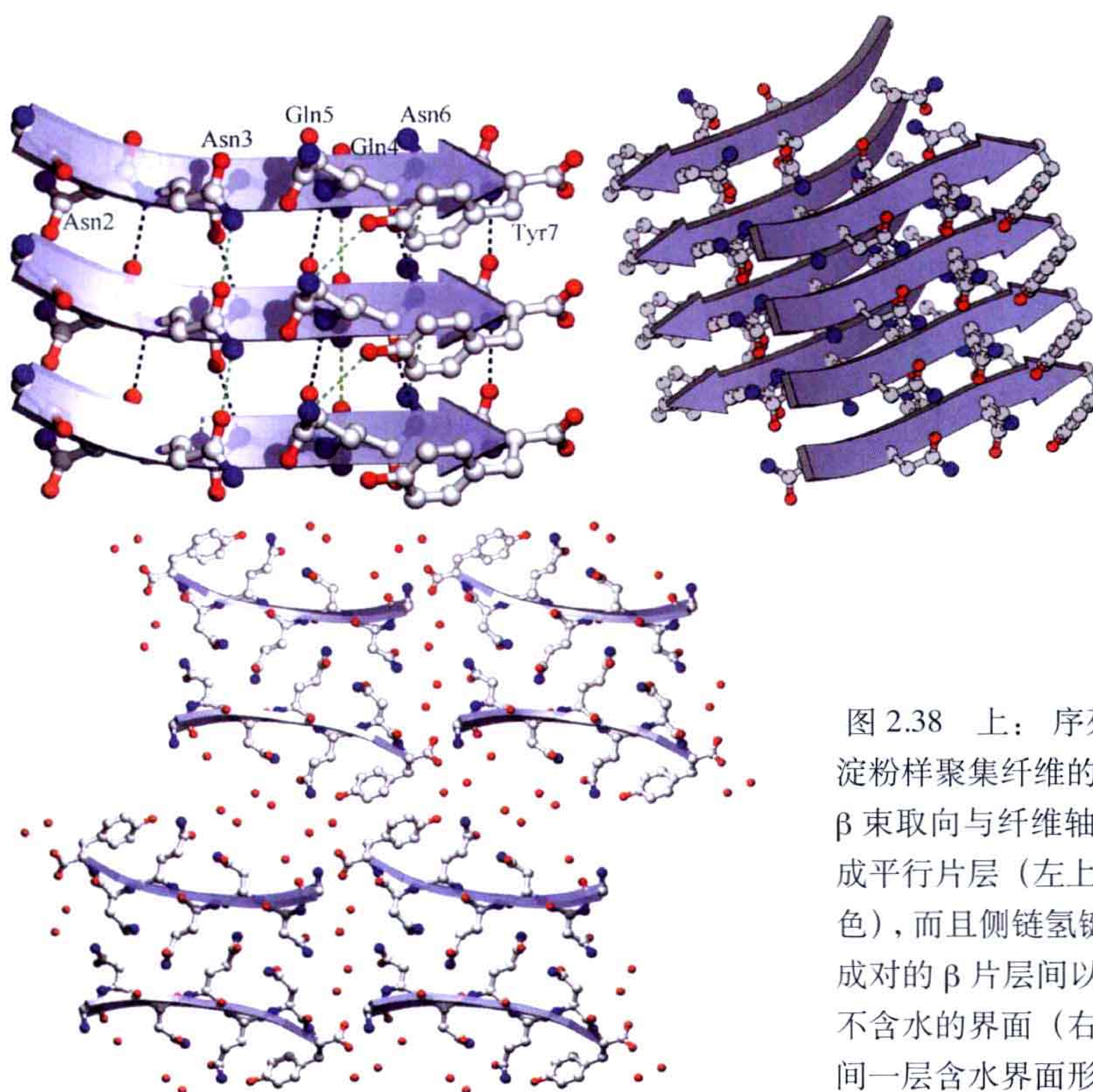


图 2.38　上：序列为 GNNQQNY 的肽形成淀粉样聚集纤维的结构细节（PDB:1YJP）。β 束取向与纤维轴相交，多肽都是对正排列成平行片层（左上图）。不但主链氢键（黑色），而且侧链氢键（绿色）也都很好地形成。成对的 β 片层间以反平行方式堆积成为完全不含水的界面（右上图）。下：β 片层对之间一层含水界面形成（水分子由红点显示）。

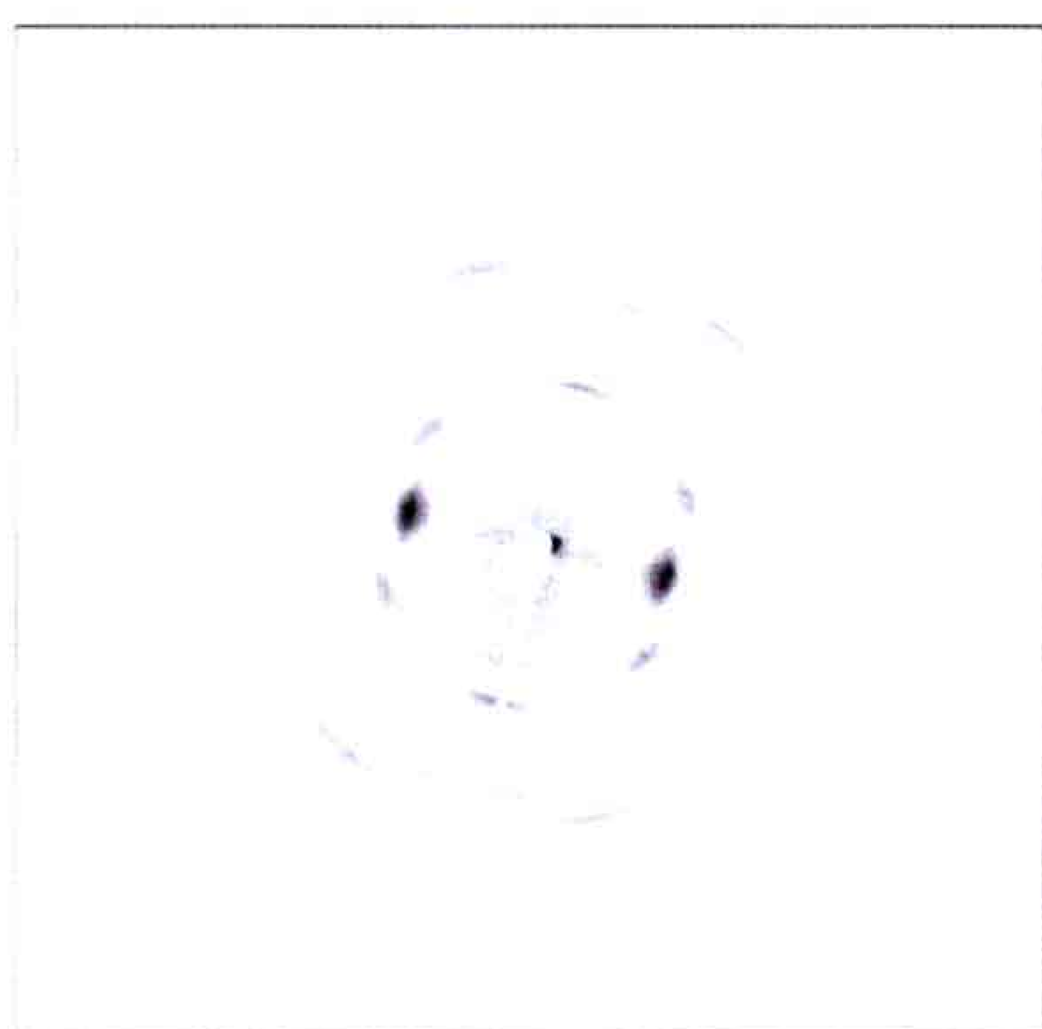
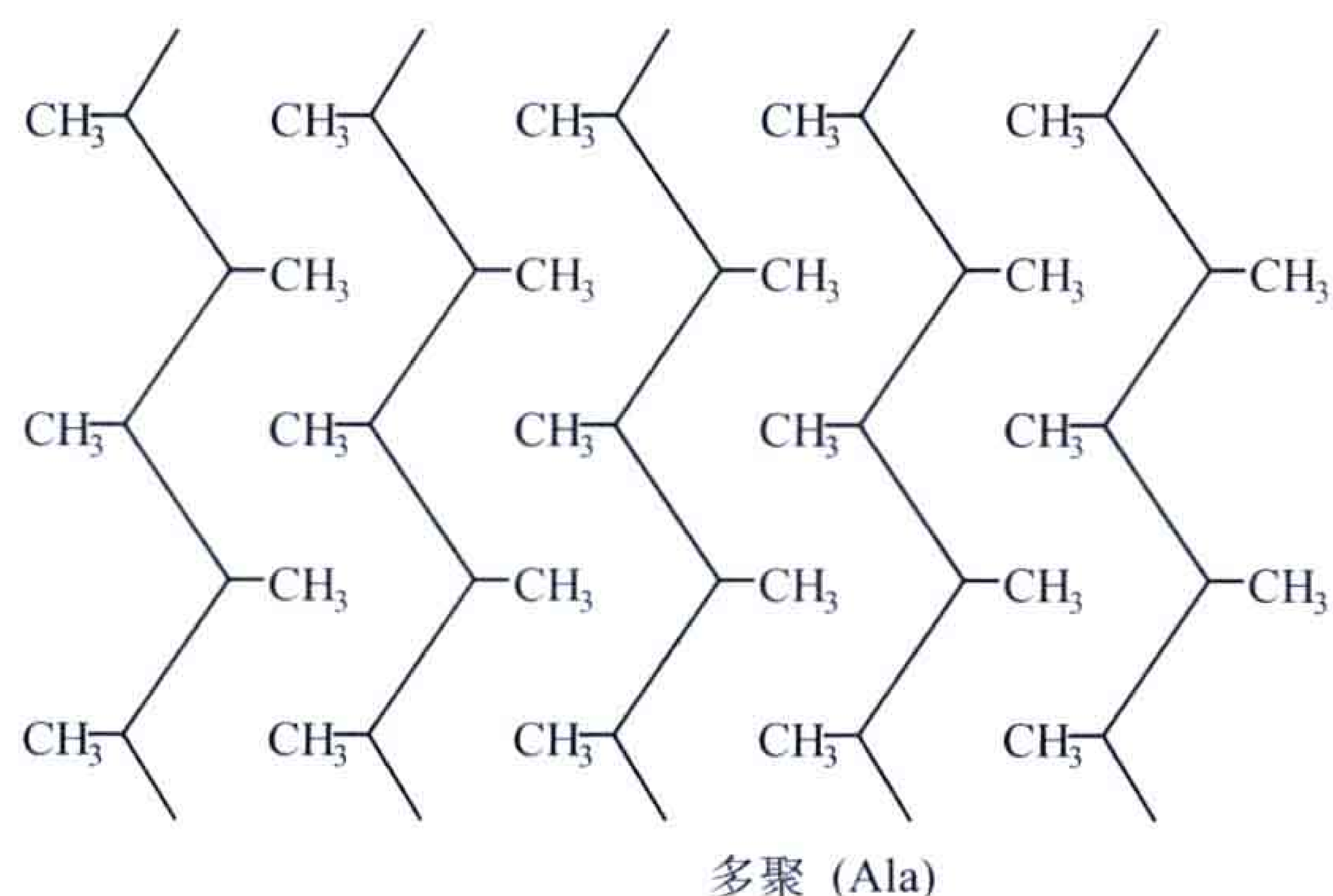
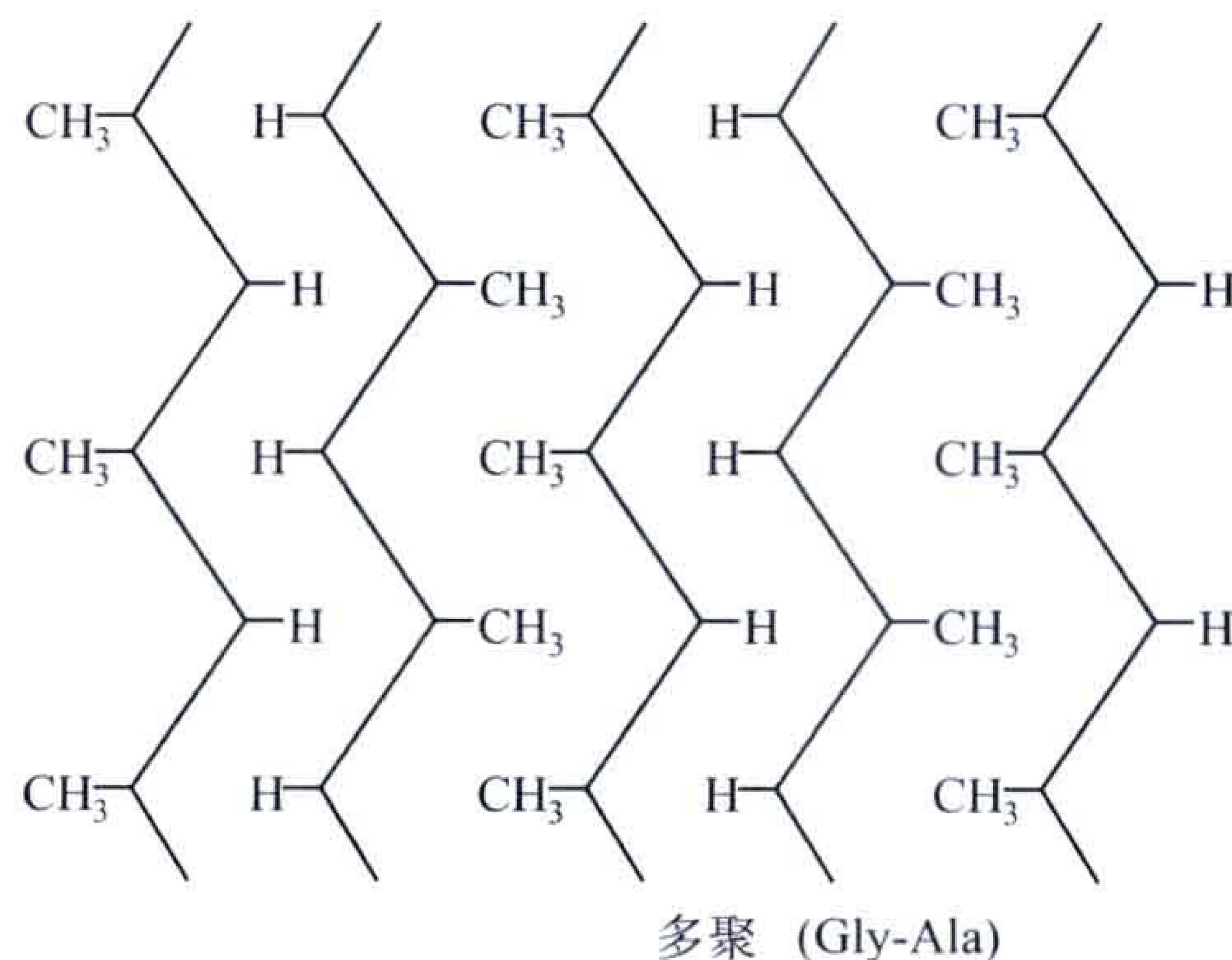


图 2.39　上：蚕丝的衍射花样 [得到许可重印自 Sangappa *et al*. (2005) Crystal structure of raw pure Mysore silk fibre based on (Ala-Gly)2-Ser-Gly peptide sequence using Linked-Atom-Least-Squares method. *J Biosci* **30:** 259-268. Copyright (2005) Indian Academy of Sciences.) 。下：蚕丝中 β 片层组织结构。每个“之”字形线表示一层从边沿看过去的 β 片层。摞在一起的 β 片层的距离取决于侧链的大小，正如甘氨酸 - 丙氨酸 (Gly-Ala) 多聚体所示。



多聚 (Ala)



多聚 (Gly-Ala)

2.2.4　蛋白质-配体相互作用

2.2.4.1　结合水

绝大部分蛋白质分子在水中折叠及行使功能。它们与水的关系是讨论蛋白质结构和功能的中心内容。两个常用到的术语是“亲水”与“疏水”，分别表示喜爱水及排斥水的性质。蛋白质表面可以是两种性质之一。很重要的一个方面是蛋白质折叠主要是疏水效应。关于蛋白质与水分子相互作用的假设之一是所谓的考兹曼（Kauzmann）疏水作用经典描述。非极性基团被水分子围绕成为笼形，当其他非极性基团相互作用时，这些笼形水分子会被释放出来。疏水作用稳定蛋白质结构的效应主要是由于有利的熵贡献。

在大分子结构研究过程中，非常多的水分子在蛋白质的表面上、孔穴中及内部被确认。晶体结构的分辨率越高，水分子的确认就越好。在第一水合壳层的水分子是蛋白质整体的一部分，直接贡献于蛋白质分子的稳定性及功能。蛋白质表面可被晶体学确认的水分子通常与一个或多个蛋白质表面的极性残基形成氢键（附录 A.3）。在某些部分，第二层水合分子也可被观察到一些。

水分子中氧原子与两个氢原子及两个自由电子对形成四面体几何构型，因此水分子可以同时成为两个氢键的供体及其他两个氢键的受体。这些氢键的供体及受体原子间距离为 2.7 ～ 3.3Å。对于低分辨率结构，确定水分子有困难。在特殊环境中（如与金属离子结合时），特定水分子的 pK_a 可被干扰，被研究系统的 pH 使得这些特定水分子去质子化形成 OH^-，或者质子化形成 H_3O^+。

2.2.4.2　内部水分子

水分子也可以被完全包埋在蛋白质分子内部。某些蛋白质具有部分或者完全穿过整个蛋白质的充满水的通道。这些通道可以是跨膜运输水分子、质子或者离子的功能性通道。有时水分子会形成广泛的氢键连接，然而形成小簇群的水分子可以在蛋白质中哪怕是弱极性或者非极性的空腔中被稳定。水分子的稳定是瞬时的，内部水分子的交换可以由核磁共振谱（NMR）或者中子衍射方法研究。

水分子可以被看做蛋白质在折叠时遗留于蛋白质中。蛋白质基团间部分极性的空腔会偏好于一个水分子而不是空洞。NMR 研究表明，这类水分子中有一些会很快地与外部水分子交换。蛋白质分子的“呼吸”运动无疑会帮助这些水分子的交换。

2.2.4.3　辅基和辅酶

很多蛋白质，特别是酶，只有在先结合一个非蛋白分子后才能够行使其生化功能。这些所谓辅助因子可以是无机或者有机分子。某些情况下这些辅助因子会与蛋白质紧密结合，如锌离子、铜离子等金属离子及血红素基团，它们被称为辅基（prosthetic

group）。其他一些基团与底物反应，需要在下一轮反应前释放出去，这些辅助因子常被称为辅酶（coenzyme），有时也被称为辅底物（cosubstrate）。例如，ATP、NAD^+（nicotinamide adenine dinucleotide）及 SAM（S-adenosyl methionine）。NADH 及 ATP 类的辅酶存在于所有生命形式的中心代谢途径中，它们都含有腺苷部分，意味着这个分子在生物进化的极早期就与催化反应有关。

2.2.4.4 金属结合

金属是很多生物系统、蛋白质及核酸的重要组成部分。金属可以连接结构中的不同部分以稳定三维结构，还可以中和结构中会互相排斥的负电荷。酶经常在其催化活性中心用到金属。生物系统中常见的金属有：Na、Mg、K、Ca、V、Mn、Fe、Co、Ni、Cu、Zn、Mo 和 W。不同的金属在蛋白质中偏好不同的配体。很多情况下无机金属簇成为蛋白质的重要成分。表 2.10 给出不同金属与不同的蛋白质配体作用的小结。

表 2.10 蛋白质中的金属：状态、配体及功能

金属	正常氧化态	配位数	常见的蛋白质配体	常见距离：金属 - 配体 /Å	主要生物学功能
钠 Na	I	6	C=O, H_2O	2.4	电荷载体，平衡渗透压
钾 K	I	8	C=O, H_2O	2.8	电荷载体，平衡渗透压
镁 Mg	II	6	Asp, Glu, Ser, Thr, C=O, H_2O	2.1	稳定结构，水解酶，异构酶
钙 Ca	II	6 或 7	Asp, Glu, Ser, Thr, C=O, H_2O	2.4	稳定结构，启动触发，电荷载体
锰 Mn	II，III	6	Asp, Glu, His, H_2O	2.2	光合作用
铁 Fe	II，III	5 或 6	His, Cys, Asp, Glu, Tyr, H_2O	2.1	稳定结构，氧化酶，氧气运输与存储，电子转移，固氮
镍 Ni	II	4 或 6	His, Cys, H_2O	Ni-S 2.2-2.4	氢化酶，水解酶
铜 Cu	I，II	3 或 4	His, Cys, Met	2.0	氧化酶，双氧运输，电子转移
锌 Zn	II	4 或 6	His, Cys, Asp, Glu, H_2O	2.0	稳定结构，水解酶
钼 Mo	–	–		–	与蝶呤配位相连
钨 W	–	–		–	与蝶呤配位相连

2.2.4.5 无机或有机金属簇

一些蛋白质，特别是酶，含有几个金属原子或离子组成的簇。其中一种是结合在半胱氨酸侧链上的 FeS 簇。可以有 2、3 或者 4 个铁离子通过无机硫原子桥结合成簇，这个铁离子簇再通过半胱氨酸残基结合在蛋白质上。这样的复合物称为 2Fe-2S、3Fe-4S 及 4Fe-4S。在后两种情况时，Fe 和 S 原子位于一个正方形的角上（图 2.40）。当只有 3 个铁原子时，正方形的一个角缺失。

一个极端的例子是铁蛋白（ferritin），其为一种铁储存蛋白。其蛋白质壳层由 24 个亚基组成 432 对称性，其内部孔洞直径约为 75 Å，可以容纳含 4000 多个铁原子的氧化铁。另外，在此复合物中心还有一些磷酸。这些水合铁氧化物的组织结构还不是完全清楚，但是矿物中水铁矿的结构似乎是很相关的。我们只是部分地了解铁蛋白的金属进入及矿化，以及利用铁时的再溶解过程。具有羧基侧链的三重轴可能参与这一过程。此蛋白质壳层的内部也主要是亲水性的。

一系列含金属的有机复合物也存在于蛋白质特别是酶中。最重要的例子是卟啉，它结合铁形成血红素或者结合镁形成叶绿素。

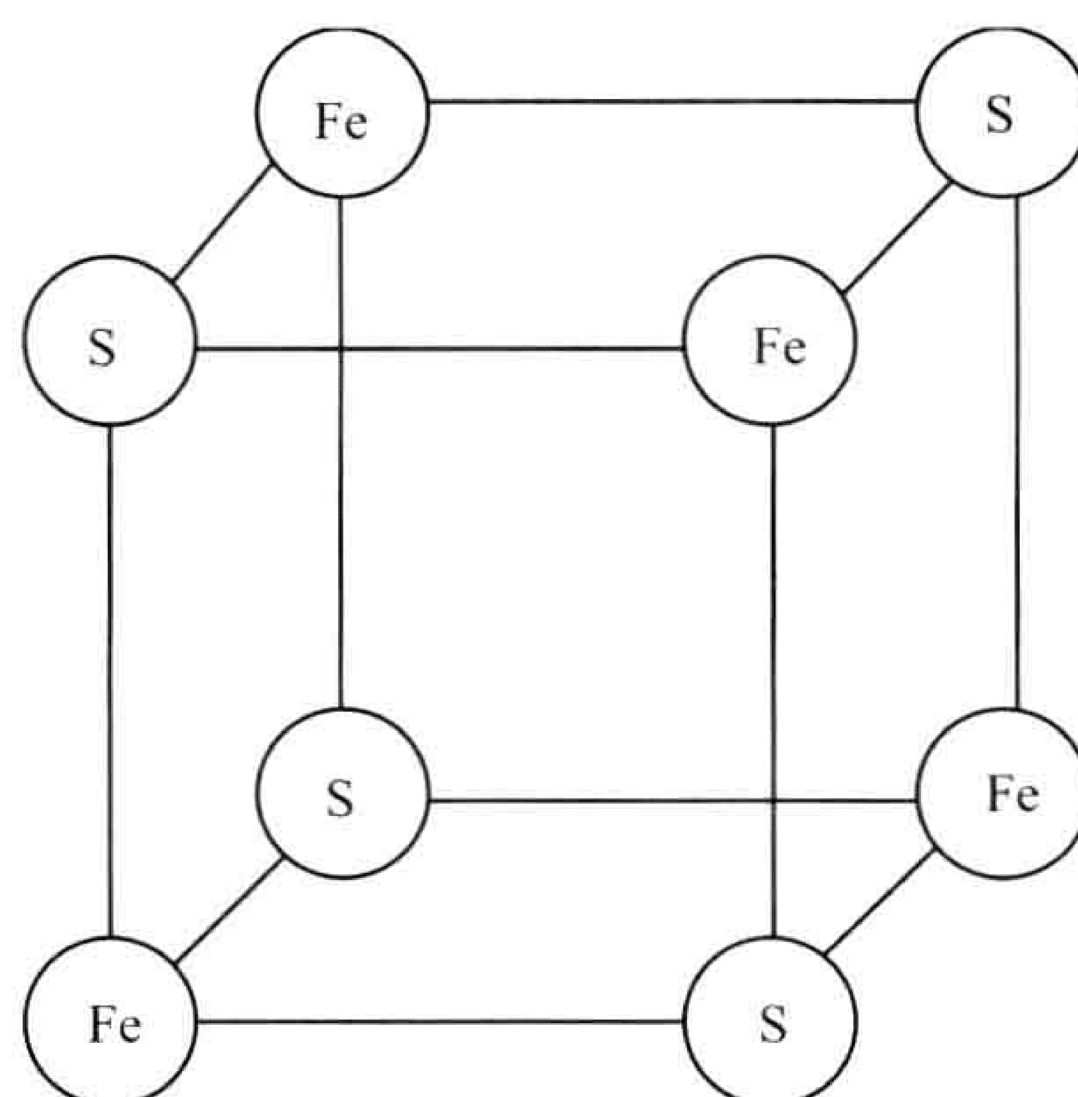


图 2.40 很多相关蛋白中的 4Fe-4S 复合物原子排列。

金属常常具有几种可利用的氧化态，并会以独特的方式参与氧化还原反应。铁、锰及铜在这方面更是独特且有趣的。

延伸阅读（2.1 节）

原始文献

Fleming PJ, Rose GD. (2005) Do all backbone polar groups in proteins form hydrogen bonds? *Prot Sci* **14:** 1911–1917.

Fodje MN, Al-Karadaghi S. (2002) Occurrence, conformational features and amino acid propensities for the α-helix. *Prot Engin* **15:** 353–358.

Lovell SC, Word JM, Richardson JS, Richardson DC. (2000) The penultimate rotamer library. *Prot Struct Funct Genet* **40:** 389–408.

Richardson JS, Getzoff ED, Richardson DC. (1978) The b bulge: A common small unit of nonrepetitive protein structure. *Proc Nat Acad Sci USA* **75:** 2574–2578.

Tompa P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* **579:** 3346–3354.

综述文章

Branden C, Tooze J. (1999) *Introduction to Protein Structure*, 2nd edn. Garland Publishing, New York.

Chothia C. (1984) Principles that determine the structure of proteins. *Annu Rev Biochem* 53:537-572.

Richardson JS. (1981) The anatomy and taxonomy of protein structure. *Adv Prot Chem* 34:167-339.

延伸阅读（2.2 节）

原始文献

Bourgeois D, Vallone B, Arcovito A, Sciara G, Schotte F, Anfinrud PA, Brunori M. (2006) Extended subnanosecond structural dynamics of myoglobin revealed by Laue crystallography. *Proc Nat Acad Sci USA* **103:** 4924–4929.

Eriksson AE, Baase WA, Zhang X-J, Heinz DW, Blaber M, Baldwin EP, Matthews BW. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*: **255:** 178–183.

Gassner NC, Baase WA, Mooers BHM, Busam RD, Weaver LH, Lindstrom JD, Quillin ML, Matthews BW. (2003) Multiple methionine substitutions are tolerated in T4 lysozyme and have coupled effects on folding and stability. *Biophys Chem* **100:** 325–340.

Lo Conte L, Chothia C, Janin J. (1999) The atomic structure of protein-protein recognition sites. *J Mol Biol* **285:** 2177–2198.

Nelson R, Sawaya MR, Balbirnie M, Madsen A , Riekel C, Grothe R, Eisenberg D. (2005) Structure of the cross-spine of amyloid fibrils. *Nature* **435:** 773–778.

Prangé T, Schiltz M, Pernot L, Colloc'h N, Longi S, Bourget W, Fourme R. (1998) Exploring sites in proteins with xenon or krypton. *Prot Struct Funct Genet* **30:** 61–73.

综述文章

Goodsell DS, Olson AJ. (2000). Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* **29:** 105–153.

Nelson R, Eisenberg D. (2007) Structural models of amyloid-like fibrils. *Adv Prot Chem* **73:** 235–282.

Ponstingl H, Kabir T, Gorse D, Thornton JM. (2005) Morphological aspects of oligomeric protein structures. *Prog Biophys Mol Biol* **89:** 9–35.

Razvi A, Schotz JM. (2006) Lessons in stability from thermophilic proteins. *Prot Sci* **15:**1569–1578.

Rose GD, Fleming PJ, Banavar JR, Maritan M. (2006) A backbone-based theory of protein folding. *Proc Nat Acad Sci USA* **103:** 16623–16633.

Soto C, Estrada L, Castilla J. (2006) Amyloids, prions and the inherent infectious nature of misfolded protein aggregates. *Trends Biochem Sci* **31:** 150–155.

Tainer JA, Roberts VA, Getzoff ED. (1991) Protein metal binding sites. *Curr Op Biotech* **2:** 582–591.

（苏晓东 译校）

第三章
核酸结构基础

细胞的遗传信息被储存在核酸中，通常是脱氧核糖核酸（DNA）。储存在序列中的信息通过翻译过程被转变为功能性的蛋白质分子，在DNA和翻译机器之间则以RNA分子作为信使来传递信息。RNA分子也可以储存信息，如在一些病毒中，但是RNA还在其他过程中扮演重要角色。跟核酸有关的处于核心地位的新陈代谢过程是复制、转录和调控，它们都依赖于蛋白质对核酸序列的识别。因此，为更好地理解核酸与蛋白质相互作用的相关生物学问题，首先理解DNA和RNA的结构信息是非常重要的。

核酸分子的骨架是由核糖（RNA中）或者脱氧核糖（DNA中）单体通过磷酸二酯键聚合成的大分子。每个核糖单体上连接一个嘌呤或者嘧啶分子，它们被称为碱基。核酸分子最重要的性质是两条聚合分子链能通过碱基对之间的氢键相互作用形成螺旋结构。大部分DNA分子很大且形成双螺旋结构（图3.1）。RNA分子由碱基配对形成的短螺旋区域和碱基与骨架通过不同方式相互作用形成的单链区域组成。

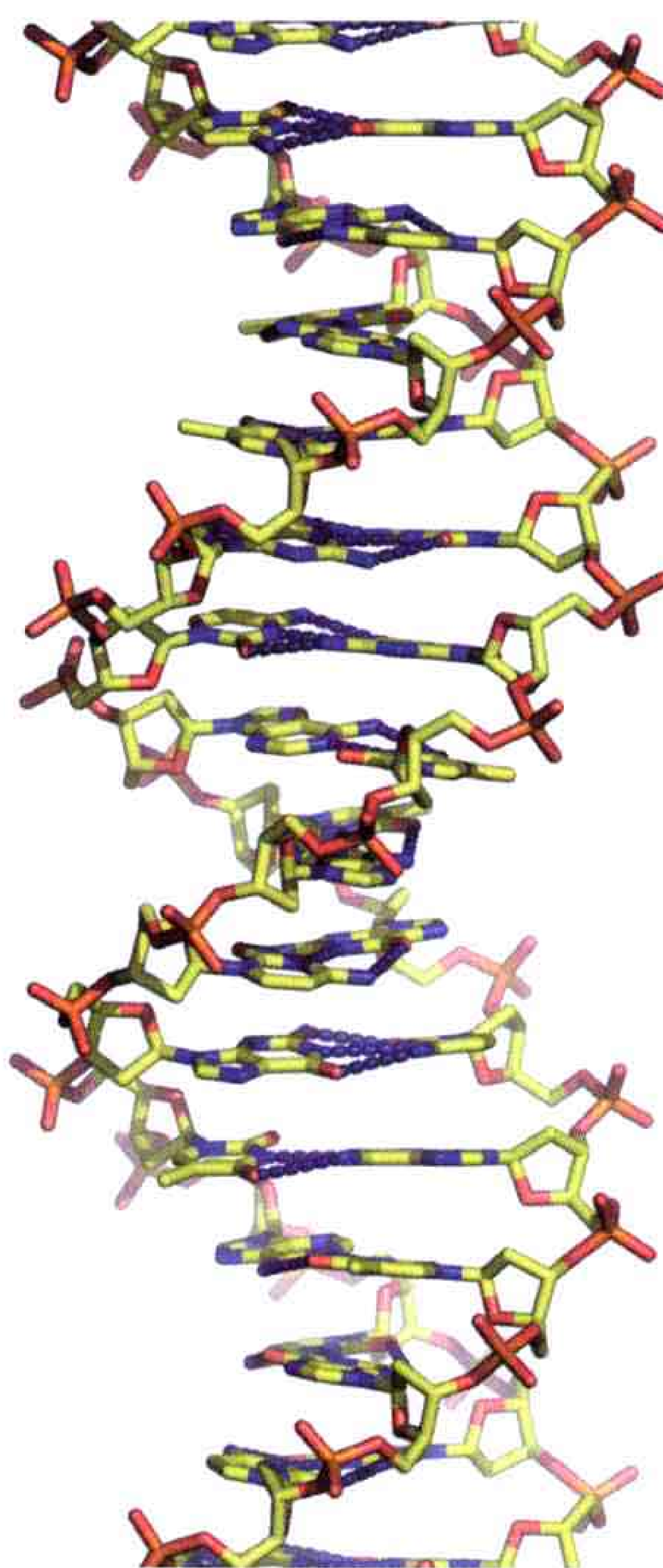


图3.1　DNA双螺旋。磷酸修饰的糖骨架以两种扭转方向（ridge）互相缠绕。碱基对在中心位置，像一摞硬币一样一个堆积到另一个上面。

3.1 构建模块

DNA 和 RNA 都是聚合物。它们的构建模块，即单体是核苷酸。核苷酸以核糖或 2′ - 脱氧核糖戊糖为骨架，在戊糖 5′ - 羟基上连有一个磷酸基团，1′ - 碳原子连接一个碱基。 在核酸聚合物中，一个核糖核酸的磷酸基团连接到下一个核糖核酸的糖基团上的 3′ - 羟基上（图 3.2）。 因此核酸分子具有 5′ 端和 3′ 端。 戊糖分子连接一个碱基后被称为核苷，在细胞中我们能找到各种各样的核苷。

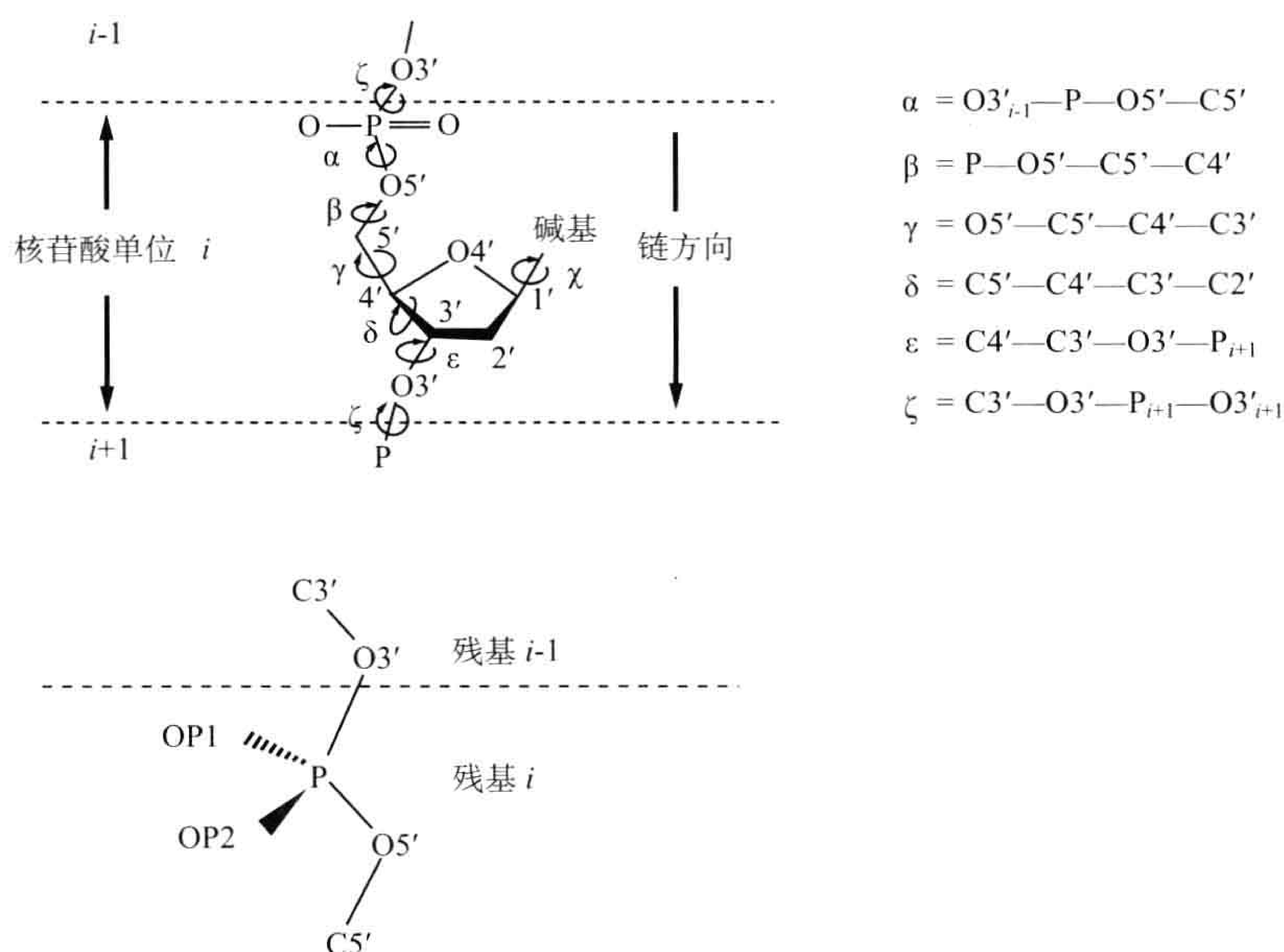


图 3.2 DNA 骨架部分，显示了原子的名称和各种扭转角的名称。其中的环是骨架的脱氧核糖部分，连续的核糖单元是通过磷酸基团连接的。因此，骨架经常被称为“糖 - 磷酸骨架”。下：链的方向的习惯性定义是通过核糖上的两个氧原子 O5′ 和 O3′ 确定的，指定与转录进行的方向相同。

骨架的构象可以用一些角度来描述，它们分别是 α、β、δ、ε 和 ζ。不过只有很少数的角度组合比较常见（图 3.3）。

在 RNA 中存在的是核糖，而 DNA 中存在的是 2′ - 脱氧核糖（图 3.4）。戊糖骨架的碳原子按顺序被编号为 1′ ~5′。核糖和 2′ - 脱氧核糖唯一的区别只有 2′ 的羟基在后者中没有出现。2′ 的羟基基团对 RNA 分子的催化性质可能有贡献。从聚合物的透视图中可以看到，3′ 和 5′ 的碳原子非常重要，因为它们直接被包含在磷酸二酯键中，把单体连接了起来。

核糖分子不是平面的而是折叠的。其有很多种可能的构象，但是最常见的构象有 C2′ -endo 构象，在 DNA 双链结构中被发现；以及 C3′ -endo 构象，在双链 RNA 中普

图 3.3　在一个分辨率为 1.96Å 的 $tRNA^{Phe}$[1EHZ] 结构中，核苷酸中各角度的相关变化。可以看到，这里有一些特定的扭转角组合是优先被采取的。Alpha-zeta 图表是最受关注的，它主要是探测磷酸原子周围的扭转角，这很容易让人联想到多肽的 Ramachandran 图。

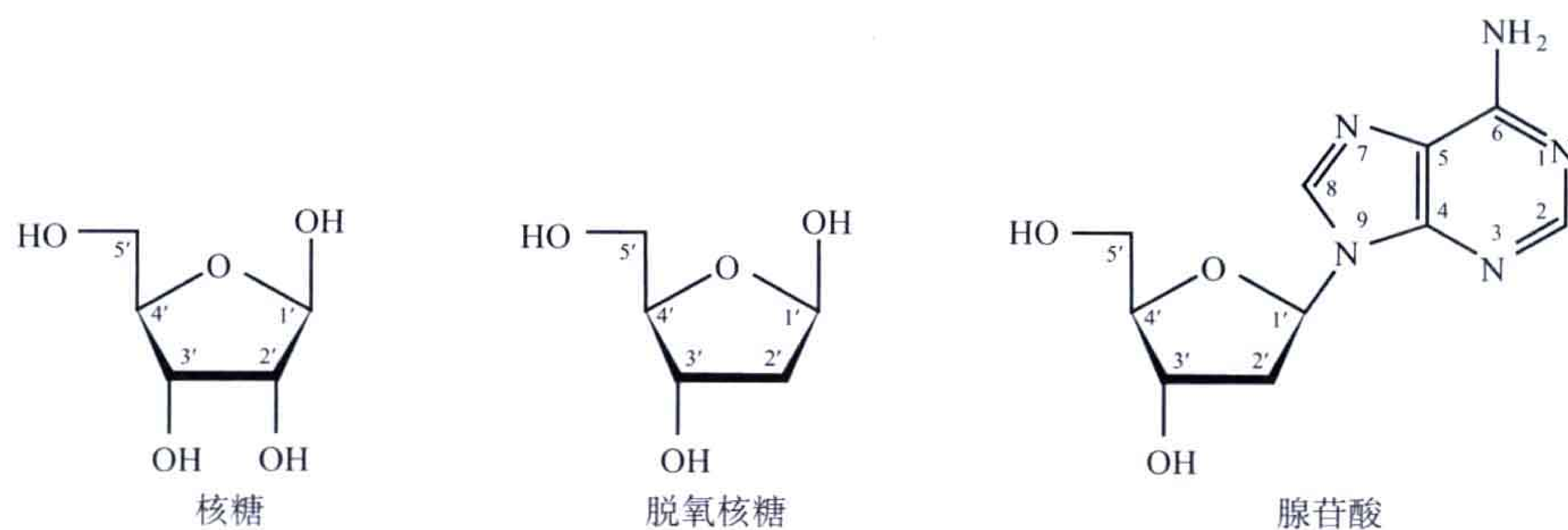


图 3.4　左、中：在核酸分子中存在的两种核糖类型；右：核苷，腺苷酸，表示了腺嘌呤碱基通过这种方式连接到了 2′ - 脱氧核糖上。

遍存在。单链 RNA 分子可能含有这两种折叠形式。所有的折叠构象可以被分为两个主要的家族，即 E 和 T。如果呋喃醛糖（aldofuranose）环的 4 个原子几乎处在一个平面上，则这种构象可以被描述为信封构象（envelope, E）（主要是画出的图像像一个信封），其余的构象可被描述为扭曲构象（twist, T）(图 3.5)。

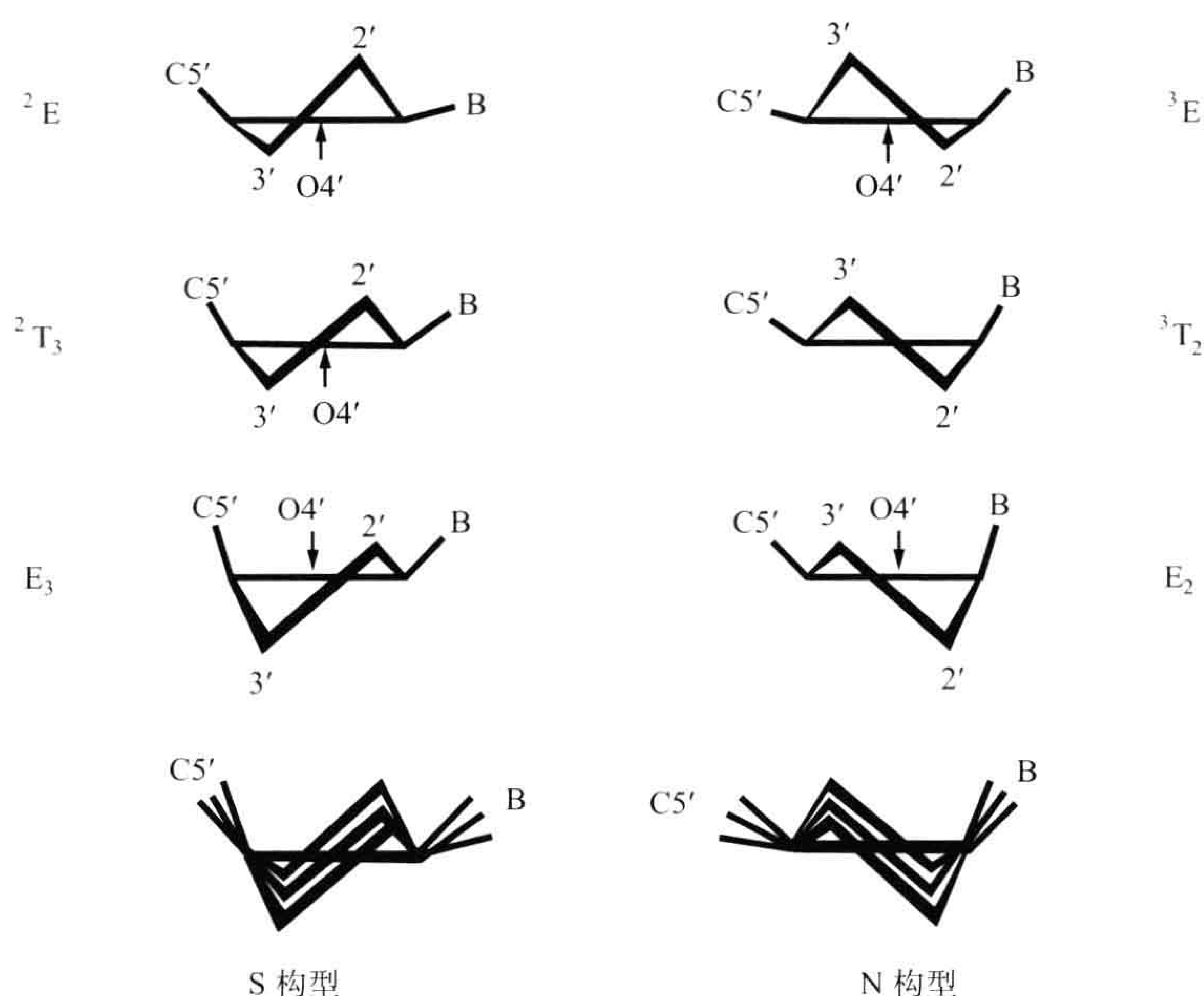


图 3.5　不同的糖环折叠构象。左边部分表示 S（south）；右边部分表示 N（north）。可以看到 C3′ -endo 构象在右上方，而 C2′ -endo 构象在左上方。E 和 T 构象的符号同样被给出。标在 E 或者 T 前上方的数字表示在同一侧的碳原子数，以 C5′ 参考平面（地平线）为界；标在 E 或者 T 后边的下标表示在参考平面另一侧的原子数。

核糖的折叠可以用相位角 P 来描述，相位角可以用内在的核糖构象角（v_0~v_4）定义为（图 3.6）

$$P = \arctan \frac{(v_2 + v_4) - (v_1 + v_3)}{2v_0[\sin(\pi/5) + \sin(2\pi/5)]}$$

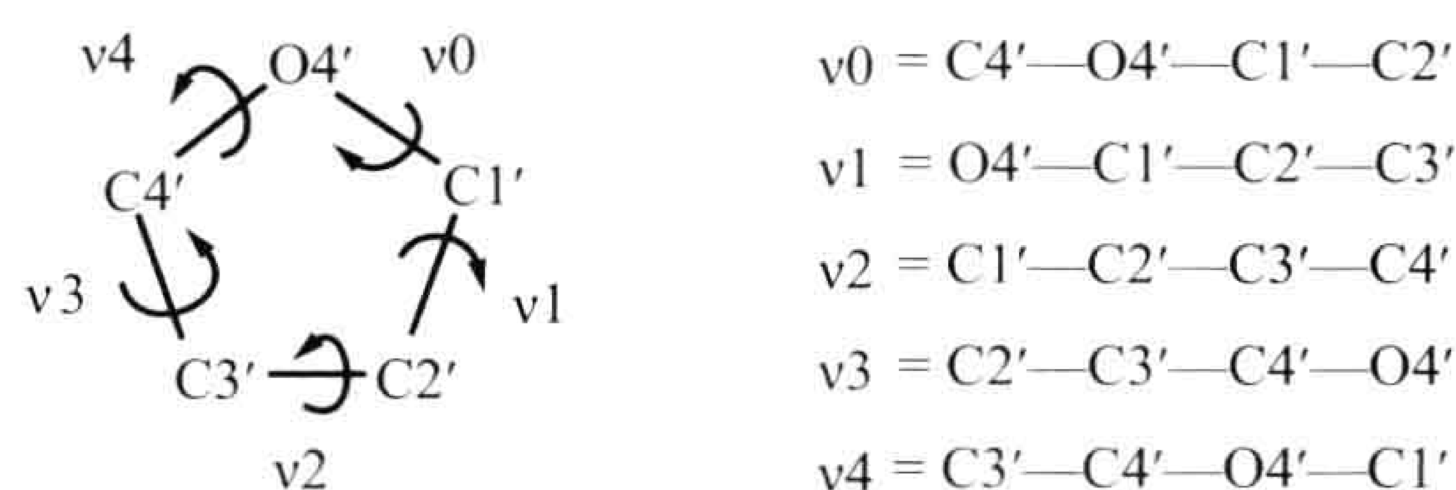


图 3.6　核苷酸中糖环扭转角命名的示意图。

P 必须为 0~360°，因此，如果 $v_0<0$，则 $P=P+180°$。核糖环不同构象的系统命名可以很方便地根据 P 环来画出（图 3.7）。

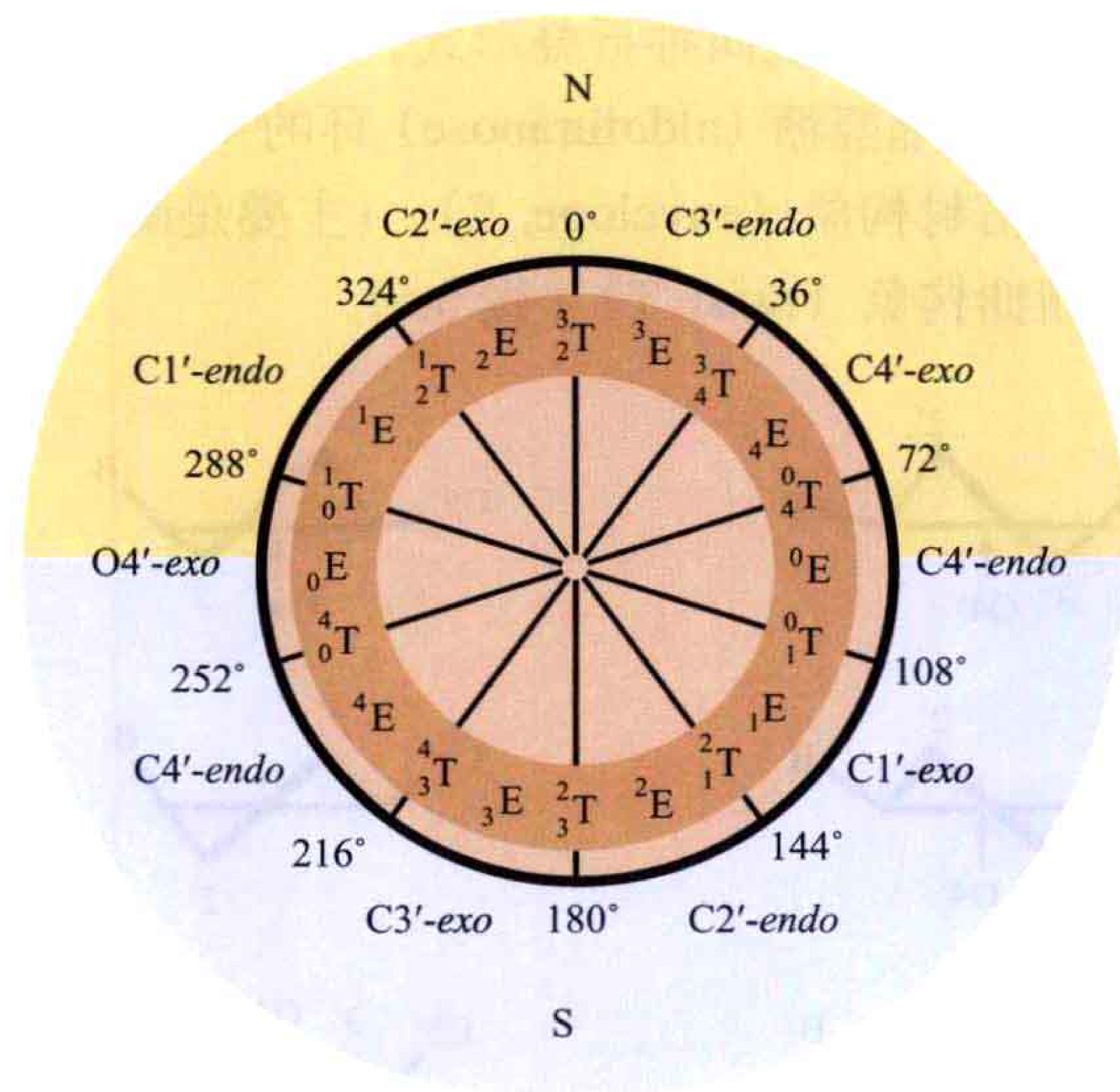


图 3.7　相位角 *P* 和核糖构象之间的关系。构象角和 *P* 被分成两个家族，north（N）和 south（S）。

在核酸分子中一共有两种最主要的嘌呤碱基类型，即腺嘌呤（adenine）和鸟嘌呤（guanine），但是存在很多的异构体。DNA 具有两类嘧啶，即胞嘧啶（cytosine）和胸腺嘧啶(thymine)。在RNA中，胸腺嘧啶通常都被一个很相似的碱基，即尿嘧啶(uracil)取代，后者相比前者只是缺少了一个甲基基团（图 3.8）。

这里有两种可取的方式可以在核糖上摆放碱基，分别是 *syn* 和 *anti*(图 3.9 和图 3.10)。在嘧啶核苷酸中，只有 *anti* 构象存在，这是因为氧原子和核糖之间的空间位阻。嘌呤可以有这两种取向，但是 *anti* 构象是常见的形式。

哺乳动物 DNA 已经被发现含有一个甲基化碱基，即 5- 甲基胞嘧啶（m5C）。除此之外，细菌 DNA 含有另外两个甲基化碱基，分别被称为 N6- 甲基腺嘌呤（m6A）和 N4- 甲基胞嘧啶（m4C）。在细菌中，m5C 和 m4C 的主要功能是保护其不被自身核酸

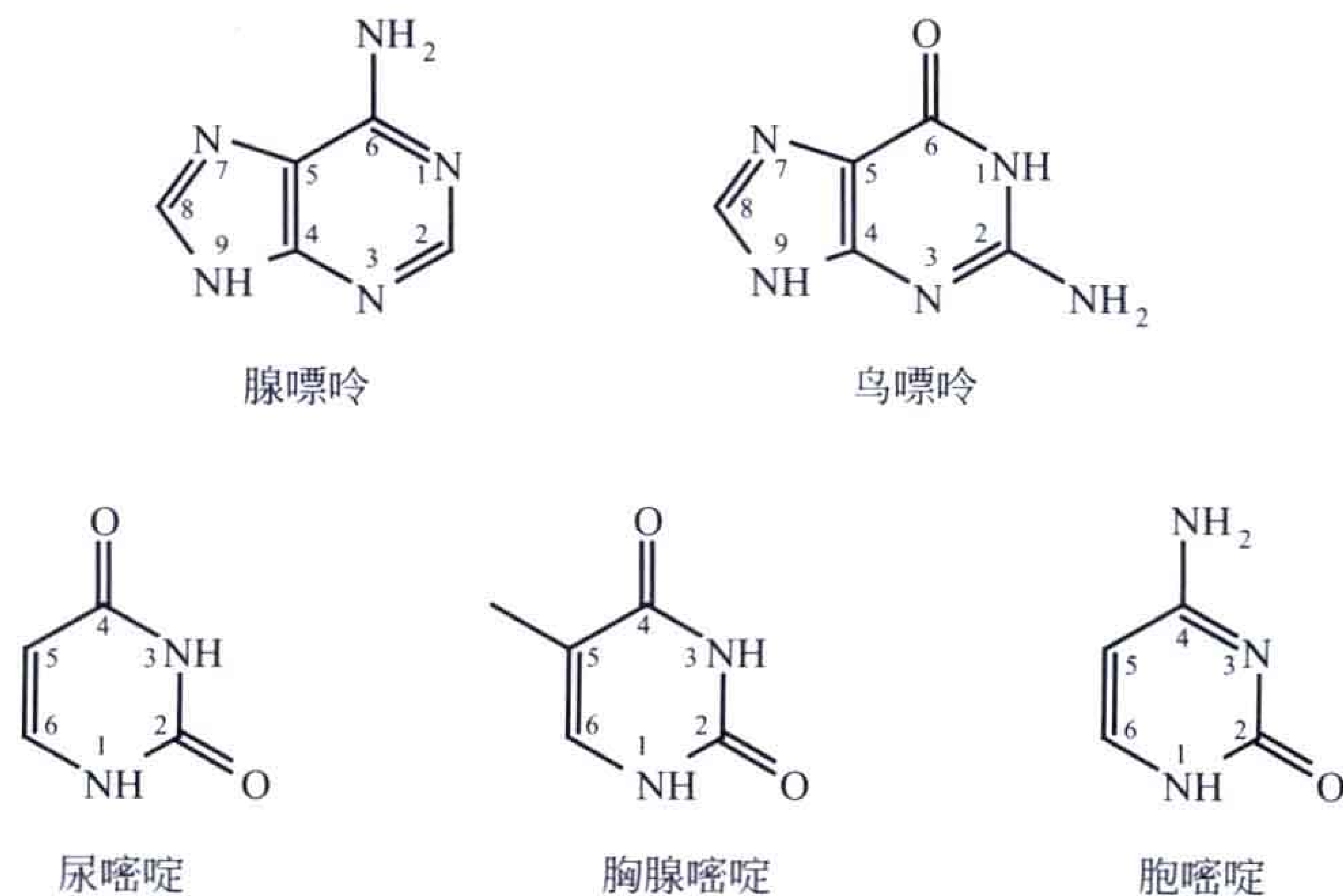


图 3.8　在核酸中存在的最常见的碱基。上：嘌呤；下：嘧啶。嘌呤和嘧啶的原子编号图也被给出。

酶的识别，这样在细胞中只有外来 DNA 和不是自己的遗传材料才会被降解。甲基化的腺嘌呤（m6A）被认为出现在致病因子的调控中，以及可控制一些细菌 DNA 的功能，如 DNA 的复制、修复、表达和转运。

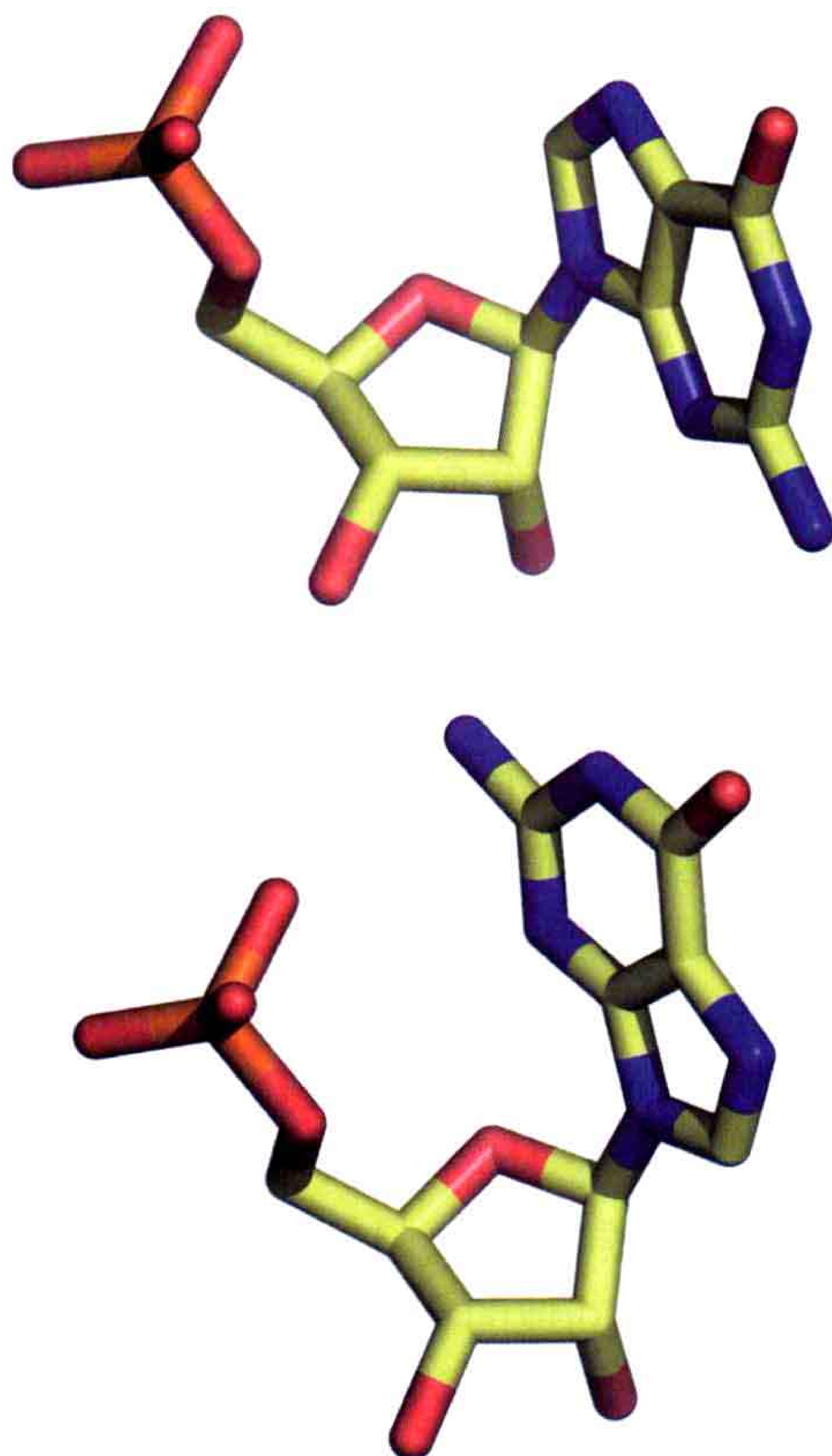


图 3.9　腺嘌呤单磷酸 *anti* 和 *syn* 构象。

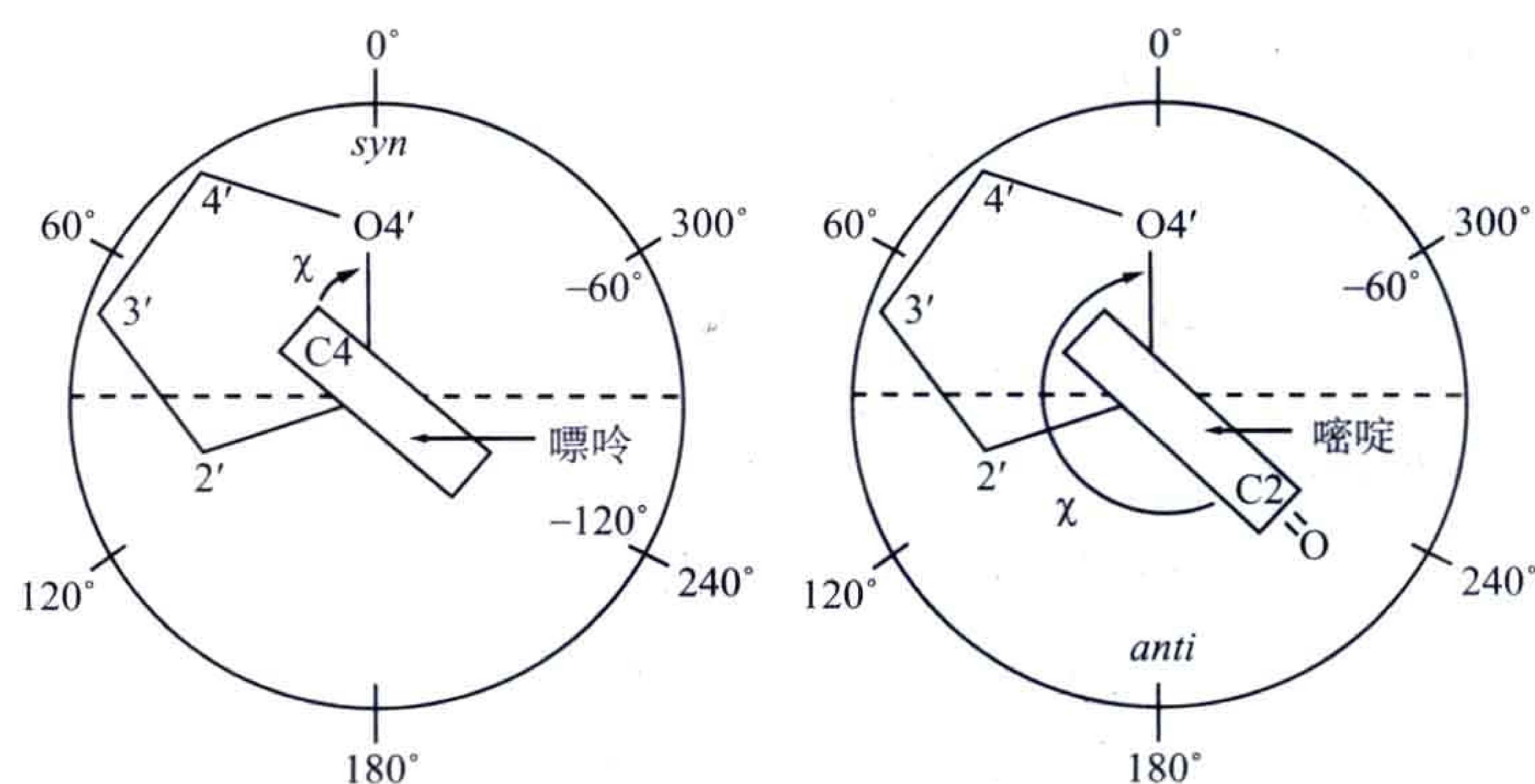


图 3.10　N- 糖苷键周围的扭转角 χ。五角形表示核糖单元，碱基在侧面。用来定义这个角的原子序列分别可以取 O4′ –C1′ –N9–C4 （嘌呤） 和 O4′ –C1′ –N1–C2（嘧啶）。因此当 χ=0° 时，O4′ -C1′ 键分别被 N9-C4 键（嘌呤）和 N1-C2（嘧啶）遮盖。*syn* 构象用 χ=±90° 定义而 *anti* 构象用 χ=180° ±90° 定义；因此，*syn* 构象区域在上半部分给出，而 *anti* 构象在下半部分给出。嘌呤在左边的 *syn* 构象区域，而嘧啶在右边的 *anti* 构象区域。

3.1.1 碱基和核苷类似物是潜在的毒性物质

碱基和核苷的类似物，结构上只是一些细微的改变，模仿了一些天然的构建模块。用化疗试剂来治疗一些类型的癌症就是基于碱基和核苷类似物的。例如，5- 氟二氧嘧啶（5-FU）成功地使用了几十年，用来抗直肠结肠癌。5-FU 需要进入细胞的代谢系统到核糖核酸水平才能体现出它的毒性。它能以 5- 氟 -2′ - 脱氧尿苷 -5′ - 三磷酸的形式被重组进 DNA，从而影响 DNA 结构。一种代谢的产物——5- 氟 -2′ - 脱氧尿苷 -5′ - 单磷酸抑制了在核酸前体新陈代谢中一个重要酶，即胸苷酸合成酶的活性，表现出非常久的效果。DNA 含有胸苷，而 RNA 没有，它是尿嘧啶的甲基化衍生物（图 3.8）。胸苷酸合成酶（thymidylate synthase）催化 dUMP 形成 *de novo*-dTMP（胸苷酸）。有两种酶，即 ThyA 和 ThyX，都能合成 dTMP，并且其具有完全不同的折叠和起源，且在不同的组织中被发现。

核苷碱基或者核糖部分被修饰的核苷类似物被广泛用于抗病毒、抗细菌和抗癌药物研制方面。一般地讲，这些类似物都是药物前体，在细胞中通过磷酸化来成为活性物质,其目的是为了达到抗增殖效应。磷酸化是通过脱氧核苷酸激酶(deoxyribonucleoside kinase）来催化实现的。因为磷酸基团的负电荷，磷酸化也同时把类似物带到了细胞内。这里有一些不同的途径，可使 DNA 合成被核苷类似物阻止。很多类似物在核糖部分会有很大的改变，经常在 2′ 或者 3′ 位出现，通过 DNA 聚合酶或者逆转录酶导致 DNA 合成在链延伸阶段被终止（图 3.11）。3′ - 叠氮 -2′，3′ - 二脱氧胸苷（AZT，zidovudine，图 3.12）就是一个例子。其他的在碱基上的替代品，如 9-β-D- 阿糖呋喃糖 -2- 氟嘌呤（fludarabine，F-araA，图 3.12），导致了 DNA 聚合酶的抑制或是功能损坏（图 3.12），如果加一些类似物，同样可以抑制 DNA 聚合中 dNTP 的供给，因此增加了替代品在聚合酶上的作用。

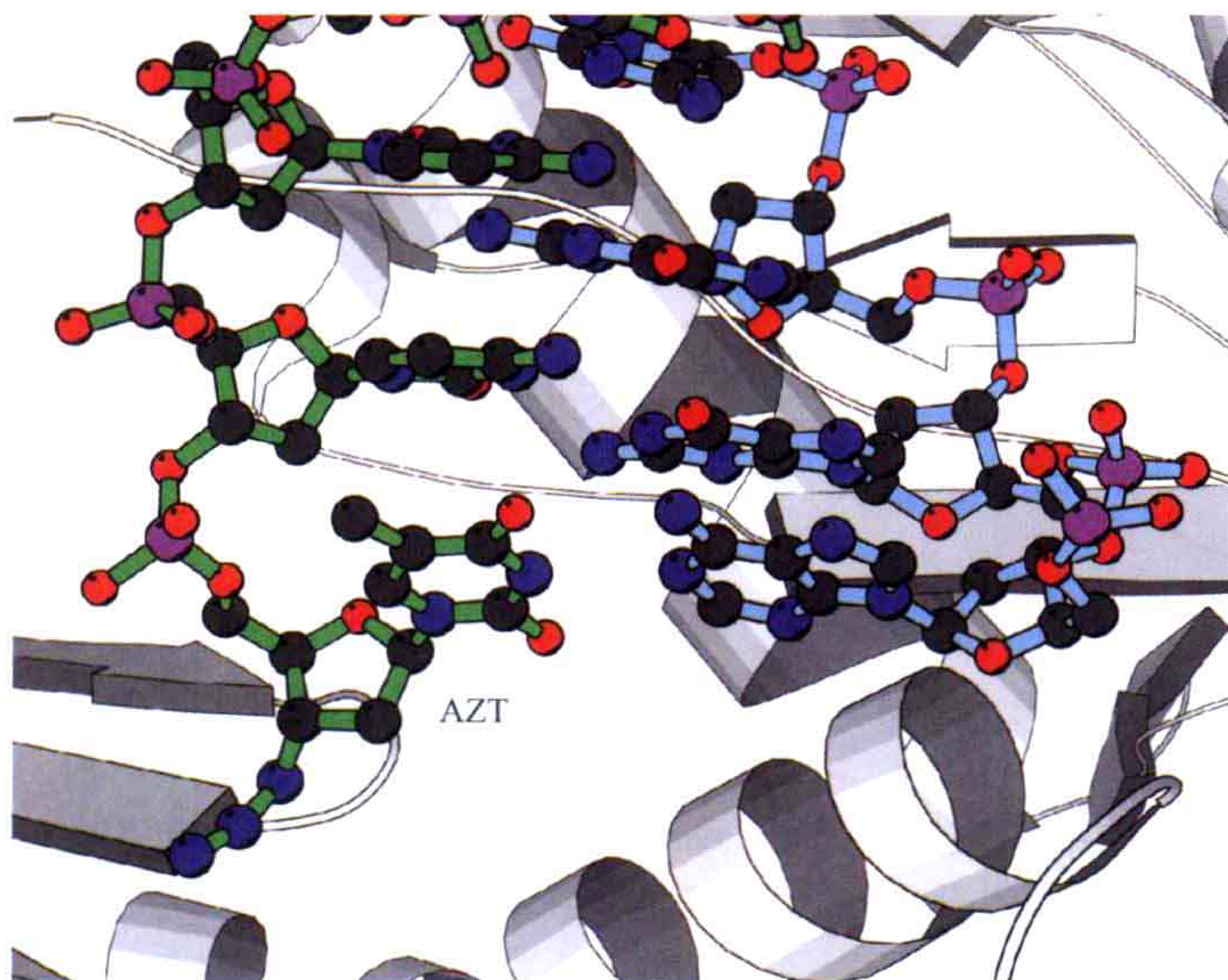


图 3.11 AZT 的作用机制：当 AZT 被引入增长的多核苷酸，其中基团 N3 防止了多聚核苷酸更进一步的增长。图表述了含有模板的 HIV-1 逆转录酶的活性位点（亮蓝色）和增长链（绿色）（PDB: 1N5Y）。

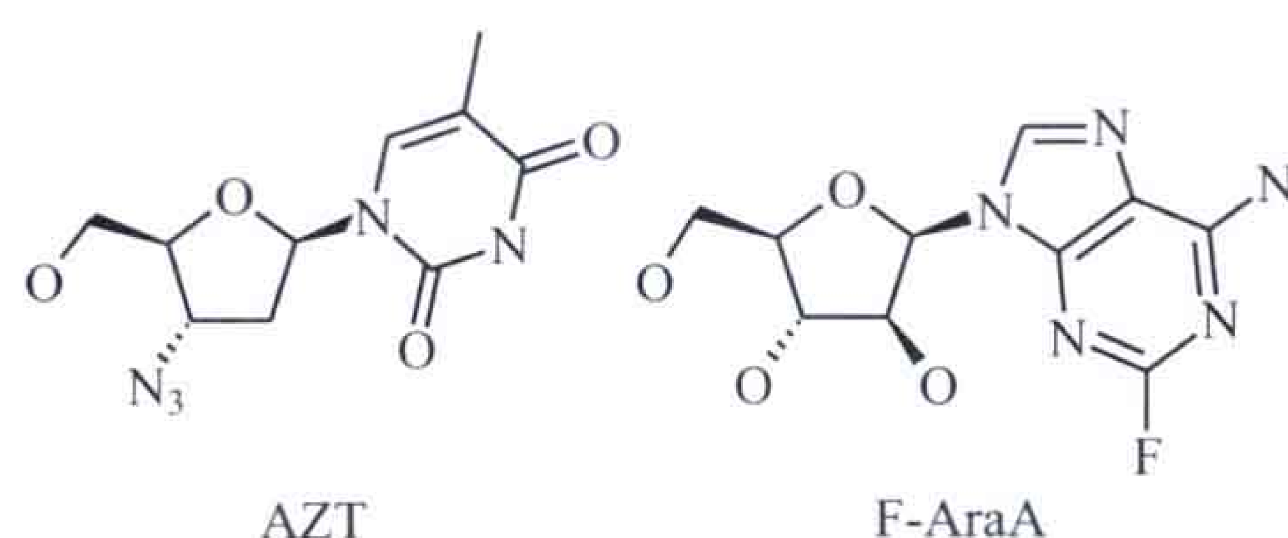


图 3.12　左：3′ - 叠氮 -2′，3′ - 脱氧胸苷（AZT，zidovudine）是第一个抗逆转录病毒胸苷类似物，被批准用来治疗 HIV 感染和 AIDS；右：9-β-D- 阿糖呋喃糖 -2- 氟嘌呤（fludarabine，F-araA）是一个成功的抗白血病药物，用来治疗 B- 细胞慢性淋巴细胞白血病（CLL）。

3.2　DNA 的二级结构：双螺旋

DNA 可以以单链或者双螺旋形式存在。核酸分子形成碱基对的能力是所有生命的基础，一条核酸链的碱基序列一定通过碱基配对和另一条链结合。一条单链在合成另一互补序列时可以用作模板。多聚核苷酸的方向可以考虑用相连的两个核苷酸之间的磷酸二酯键来定义。在核糖和磷酸部分，磷酸基团连接一个脱氧核糖部分的 3′ 碳原子和接下去的另一个脱氧核糖部分的 5′ 碳原子，因此成功地把脱氧核糖连接起来。一条链具有不同的两端：5′ 碳原子没有连接另一个核苷酸的一端被称为 5′ 端；另一端被称为 3′ 端。这两端都含有或缺少磷酸基团。

DNA 的 4 种碱基中每一个都含有唯一的氢键受体和给体形式，所以只允许和特定的碱基形成碱基对。在双链 DNA 中，我们发现 AT（腺嘌呤 - 胸腺嘧啶）碱基对含有 2 个氢键，而 GC（鸟嘌呤 - 胞嘧啶）碱基对含有 3 个氢键（图 3.13）。这些相互作用被称为 Watson-Crick 碱基对，是为了纪念第一个提出遗传基础的这两位科学家。

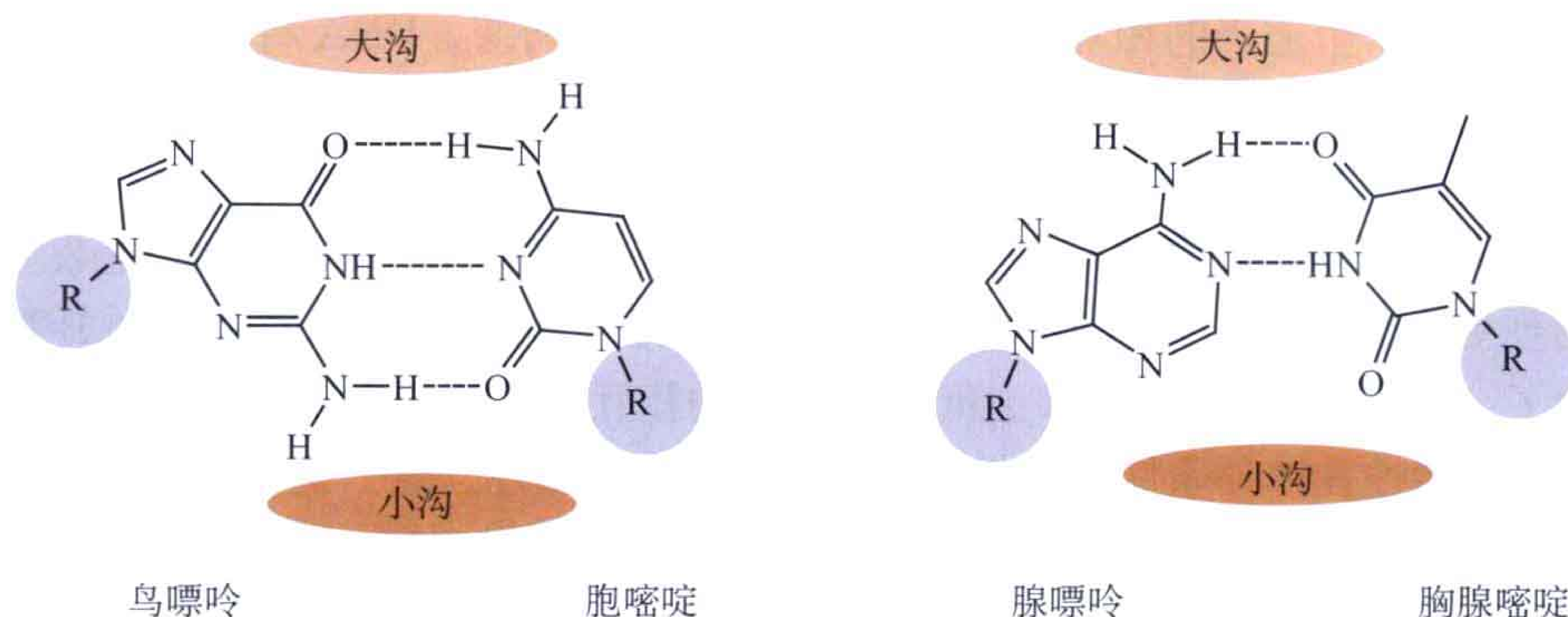


图 3.13　Watson-Crick 碱基对。核糖由 R 表示。注意，左边的 GC 碱基对通过 3 个氢键相互作用，然而右边的 AT 碱基对只有 2 个。这导致了 GC 碱基对和 GC 含量高的 DNA 比较稳定。

在双螺旋中，碱基的平面垂直于螺旋轴，且碱基对在双螺旋的中心。两条链向相反的方向延伸形成圆柱体螺旋。一条链的 5′ 端和另一条链的 3′ 端配对。这样可以用一个箭头来描述各条链，分别从 5′ 端向 3′ 端延伸。箭头指出了不同的方向，它们是反向平行的。方向参量在单链复制中是特定的。

在理想的 DNA 双螺旋中，一圈螺旋 DNA 大约含有 10 个碱基对。通常的螺旋构型是右手的，即它会向右侧扭转，就像大部分螺丝的线条。这两条反向平行链的糖磷酸骨架形成了螺旋周围的尖脊（ridge）。DNA 双螺旋上在尖脊之间由核糖磷酸骨架形成了两条凹槽。这两条凹槽通常具有不同的宽度，因此传统上我们称之为大沟和小沟。大沟比小沟宽，碱基容易被进攻。碱基对暴露的边缘含有不同氢键的给体和受体，其可以和不同的蛋白质分子相互作用。对于不同的碱基对，在大沟和小沟中暴露的边缘会有 4 种不同的给体与受体组合出现。这在 DNA 双螺旋上构建了各种各样的图谱，可被高特异性的 DNA- 蛋白质相互作用识别（见 7.2 节）。

双螺旋结构的稳定性主要取决于两个因素：在互补链之间的碱基配对，以及相邻碱基对之间的堆积，尤其是后者。真实的 DNA 分子呈现了相当大的构型变化，主要是因为真实的碱基序列、AT/GC 含量、平衡离子的出现和去稳定蛋白的存在。因为 AT 碱基对只含有两个氢键，它容易被扭曲。AT 含量高的序列在功能上很重要，一般情况下担当了 DNA 结合蛋白的结合位点。在 TATA 框结合蛋白复合物的晶体结构中，可以看到蛋白质造成了一个有意义的 DNA 弯曲（见 3.7 节）。

双链 DNA 可以采取不同的构型，也被称为形态（form）（图 3.14）。这些构型由水的运动和平衡离子的性质决定。在细胞中，DNA 的构型还受到蛋白质存在的影响。虽然各种各样的形态在体外条件下被观察到了，但是确定它们是否在体内行使生理学功能依然是很困难的。

在水溶液中，DNA 双螺旋一般以 B 型出现，每转一圈为 10 个碱基对。一个完整的螺旋在轴方向上的距离为 3.4nm。螺旋轴穿过每个碱基对的中心位置，碱基垂直于螺旋轴堆积。

在体外，一个适当的相对湿度下，将会产生结构转变，DNA 将会采取 A 型构型。这一构型通过螺旋 - 螺旋相互作用更加的浓缩和稳定。如果裸露的 DNA 紧密堆积在一起的情况出现，大家可以想象这一构型可以在体内被找到，如在病毒中。在 A-DNA 中，碱基对不再垂直于螺旋轴而有 13° ~19° 的倾斜。螺旋轴将从碱基对的中心发生位移，将会出现更大的大沟，同时小沟出现了收缩。因此，碱基对的暴露图像和 B 型中的已经不同了。这样蛋白质和 DNA 之间序列特异的相互作用将会出现特征改变。

当水的活性在体外条件下变得很低时，C 型将会出现。这一构型压缩性较低，透明度也较低，在变性条件下也能出现。除此之外，另外一些 DNA 构型也曾被报道，主要是碱基序列和环境的原因。其中一个最特别的构型是一段 G 和 C 碱基交替的短 DNA。这一构型被称为 Z-DNA，为左手螺旋，通过二核苷酸单位的重复，每一个螺旋圈含有 6 个二核苷酸，展示了特有的“Z”字形骨架（图 3.14）。

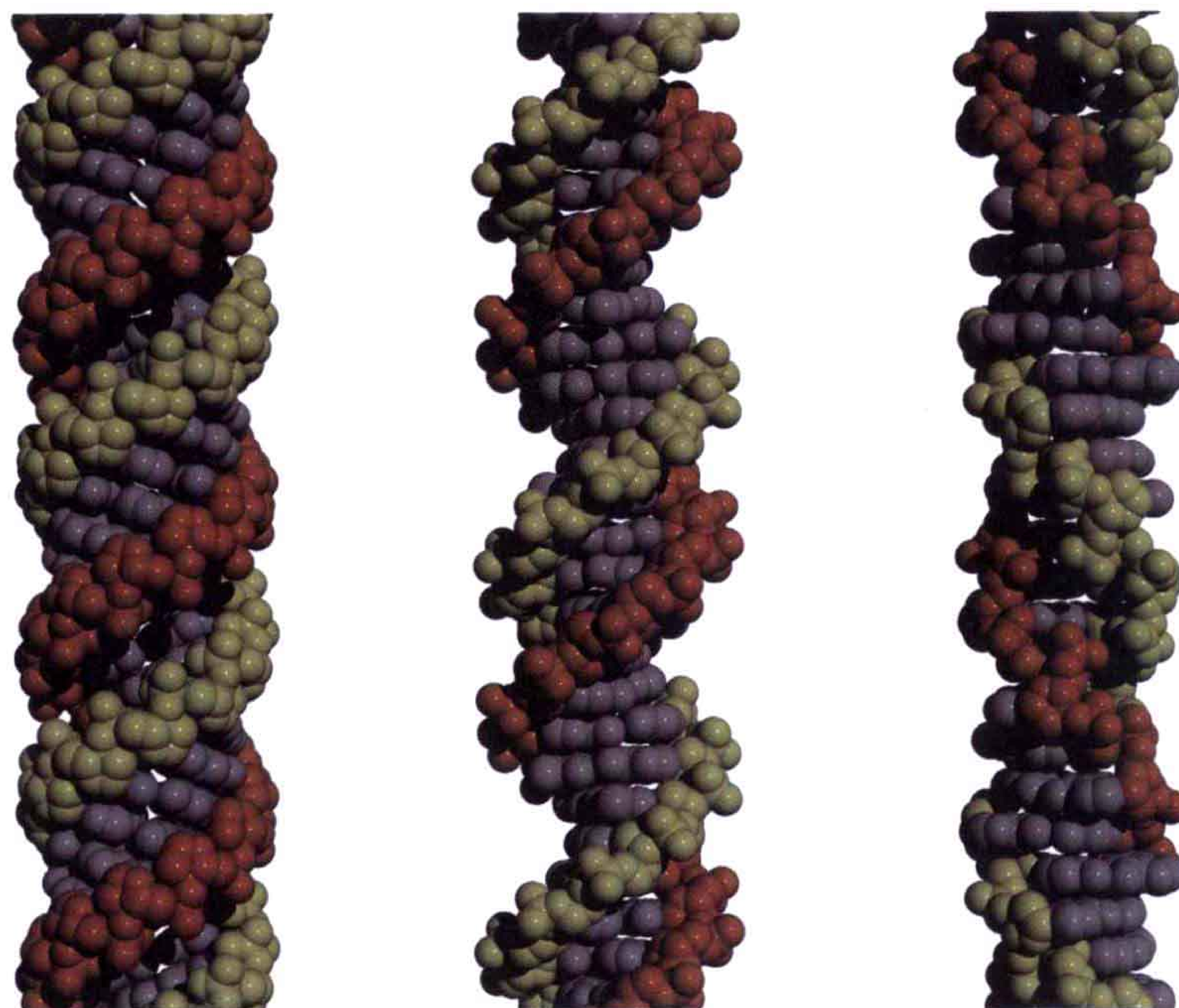


图 3.14　A-、B- 和 Z-DNA 的 CPA 模型表示。

一些最近被解出的结构显示了各种不同形状的组合，就像细胞中真实状况的反映。最近得到了一些在生物学上意义重大的结构数据，如 X 型 Holliday 交叉结构，为解释生物学过程提供了非常多的信息。这一构象是在 DNA 进行重组时得到的，两条双螺旋的 DNA 通过一个十字交叉位点连接到了一起。在这一过程中，两条螺旋部分区域交换了 DNA 链，揭示了生物变异的分子基础（图 3.15）。

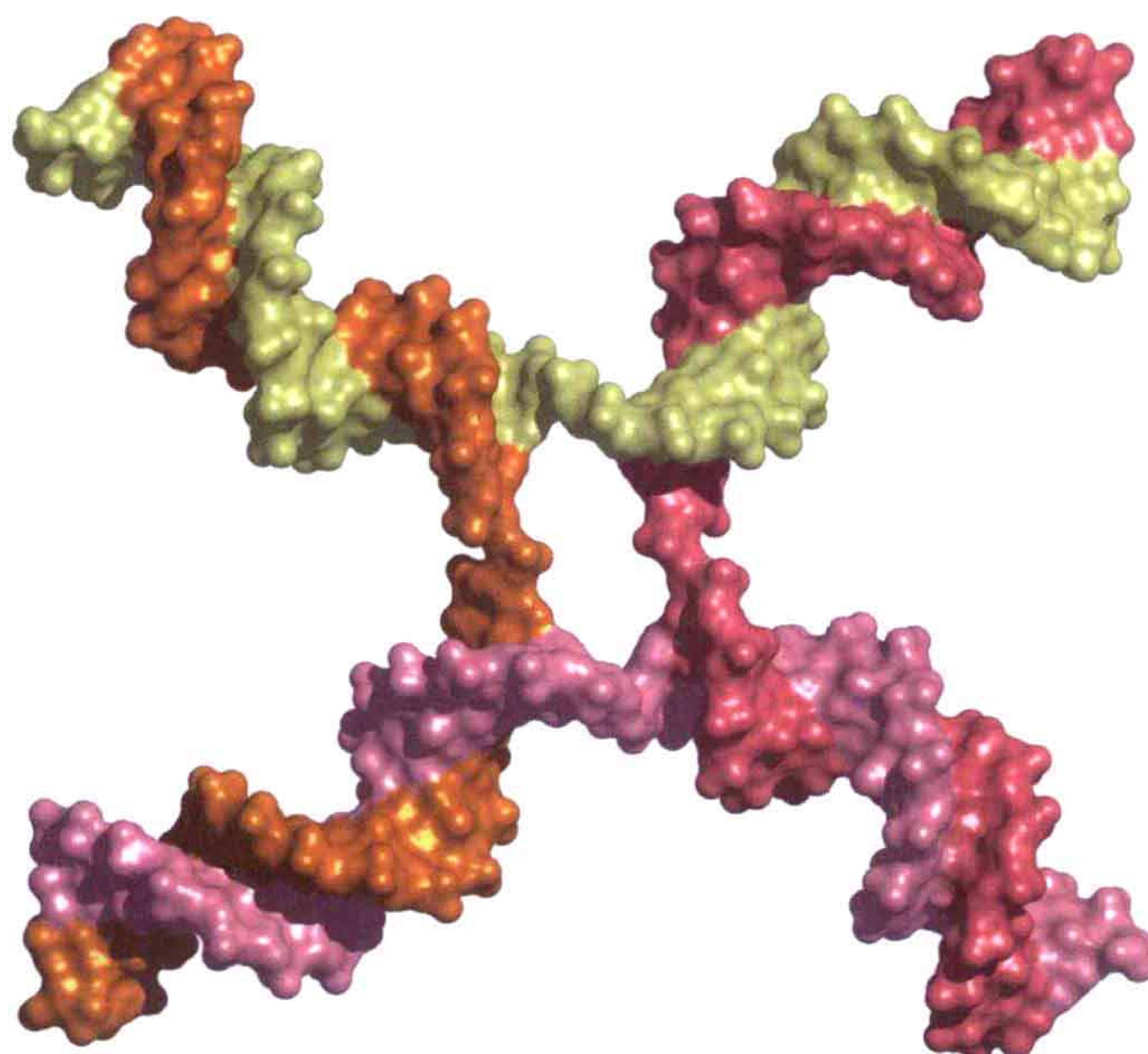


图 3.15　Holliday 连接（PDB：3CRX）。

结构生物学中的基本问题是序列如何确定构象。对于蛋白质和 RNA 来说这是非常复杂的，因为它们可以采取各种不同的构象。而 DNA 比较确定，一般采取非常有限的几种结构。人们一直希望，如果可以确定一条短链的所有可能结构，就可以把这些模型组合到一起得到一条序列的整个结构。这可以帮助人们去预测一个较长分子的结构。尽管如此，往往真实的结构要比从小模型开始预测的结构复杂得多。

DNA 在水溶液中，室温下一般采取螺旋结构。当温度升高时，螺旋将经受一个螺旋 - 卷曲转变的过程，DNA 分子将变成无规律的卷曲构型（图 3.16）。这一过程被称为变性，同时是可逆的。一个螺旋 - 卷曲转变过程可以是协同的也可以是不协同的。在完全协同的情况下，有一个全或无的特征。换句话说，任何给定的分子要么是完全螺旋状态的，要么是完全无规律卷曲状态的。变性这一性质的研究对 DNA 在生物化学方面的理解是必不可少的。简单地说，在细胞中，可逆和局部变性是 DNA 复制与基因表达中必不可少的过程。

这些构型的差异都是 DNA 分子中碱基理想堆积下得到的结果。

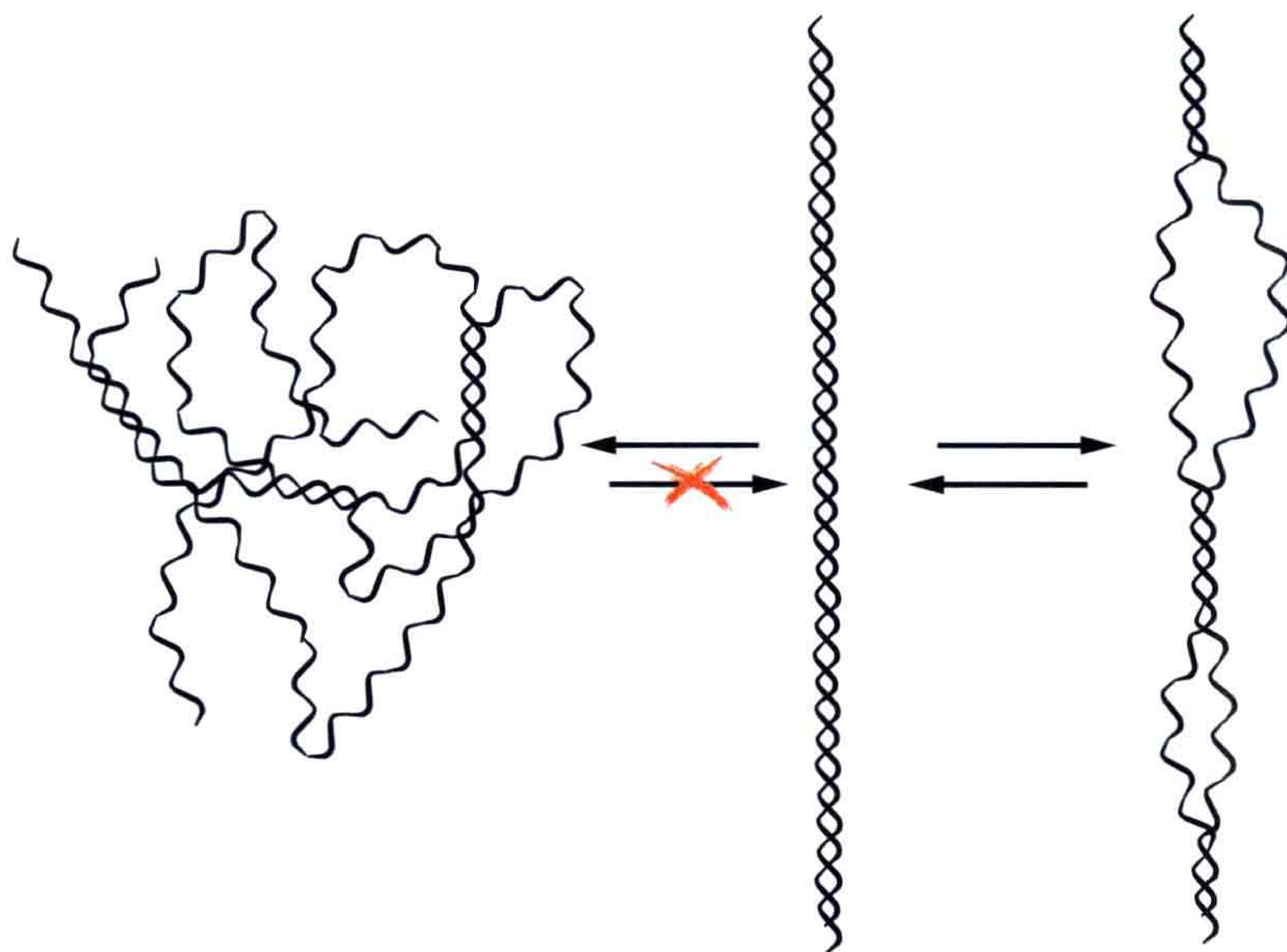


图 3.16　从双螺旋结构到局部变性转变的过程。这是一个可逆过程，但完全或是大部分的变性是不可逆的。在完全变性的情况下，随机的碱基配对可能出现，因为一直会有大量的短互补序列。

3.2.1　理想结构的偏差

DNA 双螺旋结构不是完美整齐的结构，也不是一个完美的棒状结构。相反，在染色体中的 DNA 采取紧紧缠绕和非常紧凑排列的结构。用来规定在双螺旋结构中碱基构象的参数已经在剑桥会议（Cambridge Convention）中被 R. E. Dickerson 明确，这些参数没有把骨架的构象考虑在内，只是考虑了相关的嘌呤和嘧啶部分的空间取向。关于

中间碱基的 6 个参数（移动、滑动、升起、倾斜、滚动、扭转）描述了沿着骨架的两个相邻碱基的部分构象。在这些参数中，前面三个参数是平移参量，后面三个是转动参量。在实际运用中，内部碱基的构象将会被表示成这 6 个参数的线性组合。

碱基对是相反链上不同碱基配对形成的，在 B-DNA 中，碱基对不是完全平面的。碱基对相对于理想平面的偏离取决于参与形成碱基对的碱基的不同。例如，A:T 碱基对具有一个典型的 −15° ~−20° 的螺旋桨式扭曲。在剑桥会议上还定义了 6 个参数用来描述双螺旋中的碱基偏离理想平面的情况（剪切、拉伸、错开、弯曲、螺旋、打开）。这些参数描述了参与形成碱基对的碱基的平移和转动情况。这些参数都是定义在螺旋轴上的标准坐标系中的（图 3.17）。

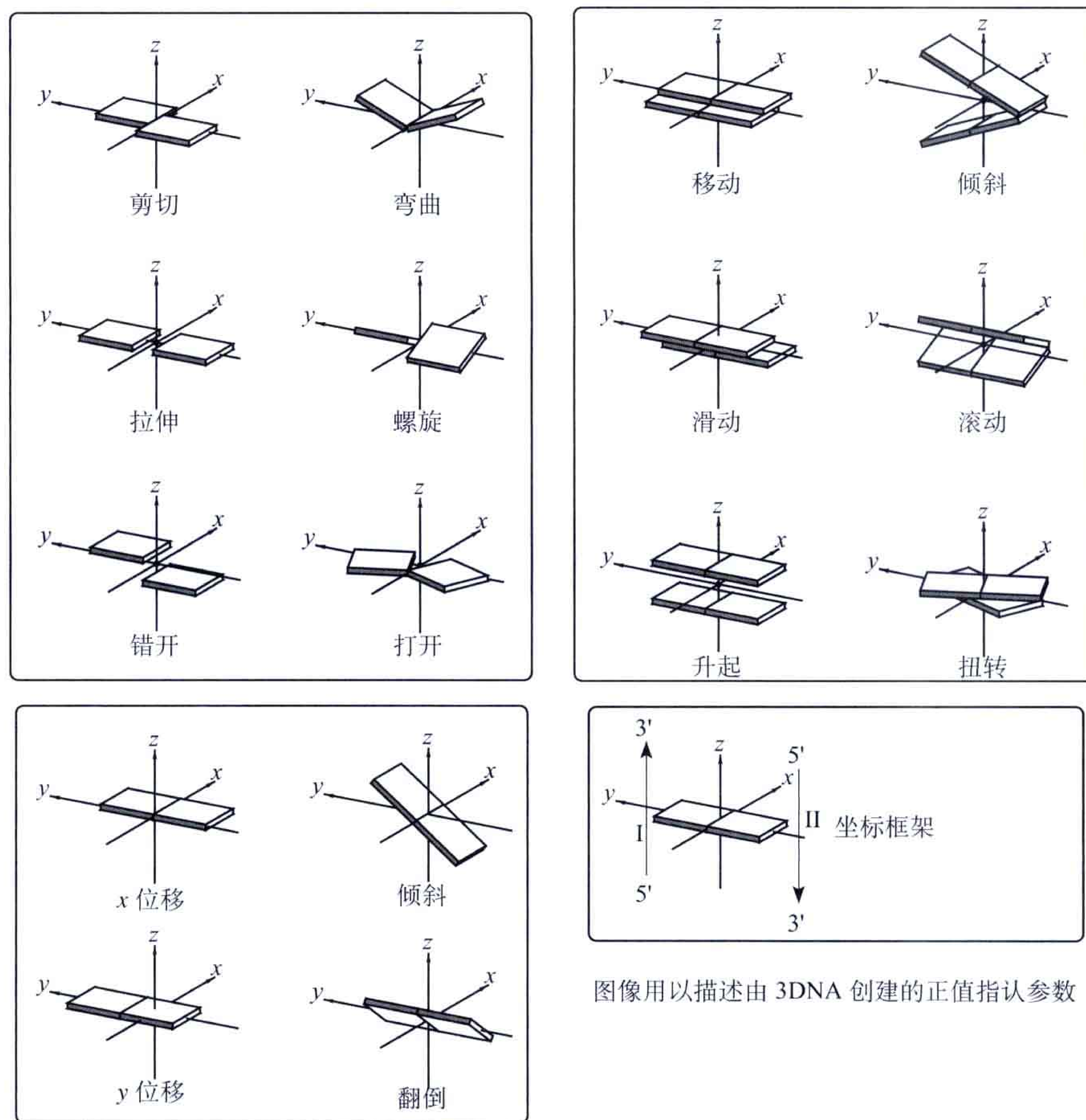


图 3.17　剑桥会议（Cambridge Convention）定义的标准参数。用来描述双螺旋中碱基的相对取向（Reprinted with permission from http://rutchem.rutgers.edu/~xiangjun/3DNA/.）。

当由具有不同取值内在碱基对参数的碱基对来搭建一个双螺旋时，螺旋的轨迹将取决于这些取值。图 3.18 显示了由于两个参数的变化而出现了不同的双螺旋构象。双螺旋将采取碱基之间范德瓦耳斯相互作用最强的构象。一些参数的组合因为空间位阻是不可能出现的，而另外一些将造成螺旋的弯曲和回转。这也受到 DNA 中特定序列的

影响。基因编码是简并的，因此 DNA 的碱基组合发生改变时可能不导致表达蛋白的突变。所以，DNA 碱基序列不仅仅是基因信息的存储器，同时也是其自身三维结构形成的原因。

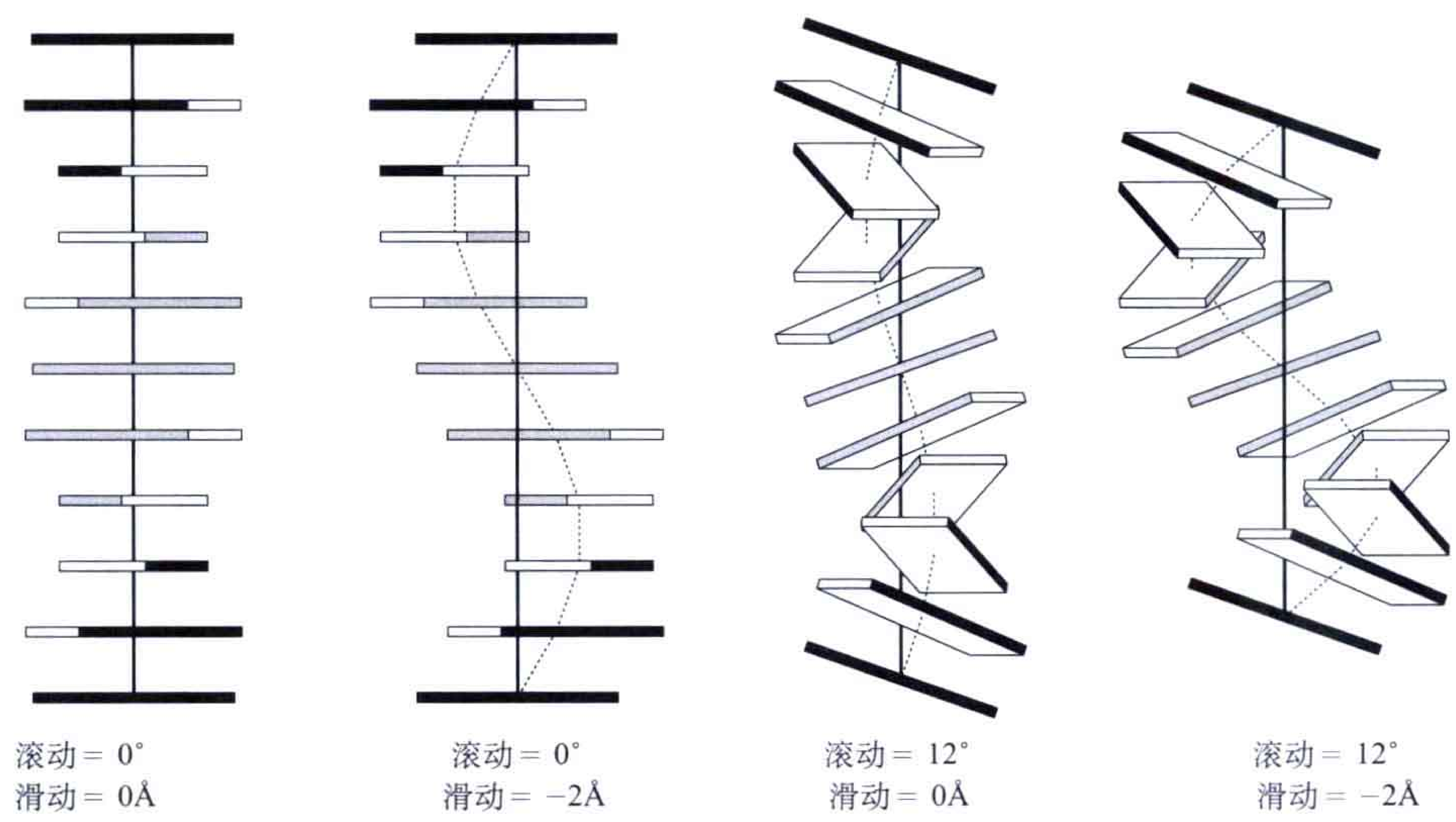


图 3.18 内部碱基参数对螺旋轨迹的影响（Reprinted with permission from http://rutchem.rutgers.edu/~xiangjun/3DNA/）。

3.3 高级结构与核小体

在细菌中 DNA 一般都被描述成裸露的分子，但是我们知道一些结构蛋白会帮助其调节它们的二级结构，同时决定三级或更高级的结构。然而这些蛋白质中的一些只是帮助中和了磷酸基团的负电荷，从而帮助 DNA 分子可以排列成更加紧密的结构，其他的一些蛋白质通过局部打开双螺旋结构来主动地转变 DNA 结构。当 DNA 蜗牛酶（DNA helicase，解旋酶）（图 3.19）结合到 DNA 的一条单链上时水解 ATP。ATP 的水解改变了解旋酶结构，因此促使它完成了机械作业（见 6.1.1 节和 7.5 节）。解旋酶驱使其自身通过分开双链的方式穿过螺旋，使局部的变性 DNA 暴露出来给其他的一些蛋白质，如 DNA 聚合酶。这样的过程在真核生物中还要复杂得多，它们的大部分 DNA 都被定位在细胞核中，同时还缠绕成特定的结构单元，被称为核小体。一条人类的染色体包含了数以百万计的核苷碱基对，假设相当的 DNA 分子被拉伸，它可以延伸到几毫米至几厘米。尽管如此，DNA 还是紧密地缠绕成只有数微米的结构。染色体的形成是动态平衡的。在染色体和细胞分裂过程中，染色体处于最紧密的状态。在细胞周期间隙，所有的压缩比例依然有上千倍。然而，染色体不同的区域浓缩到松散的结构之间具有不同的比例。压缩不仅仅对于形成结构来说很重要，同时对复制调控、DNA 修复和基因表达来说也很重要。

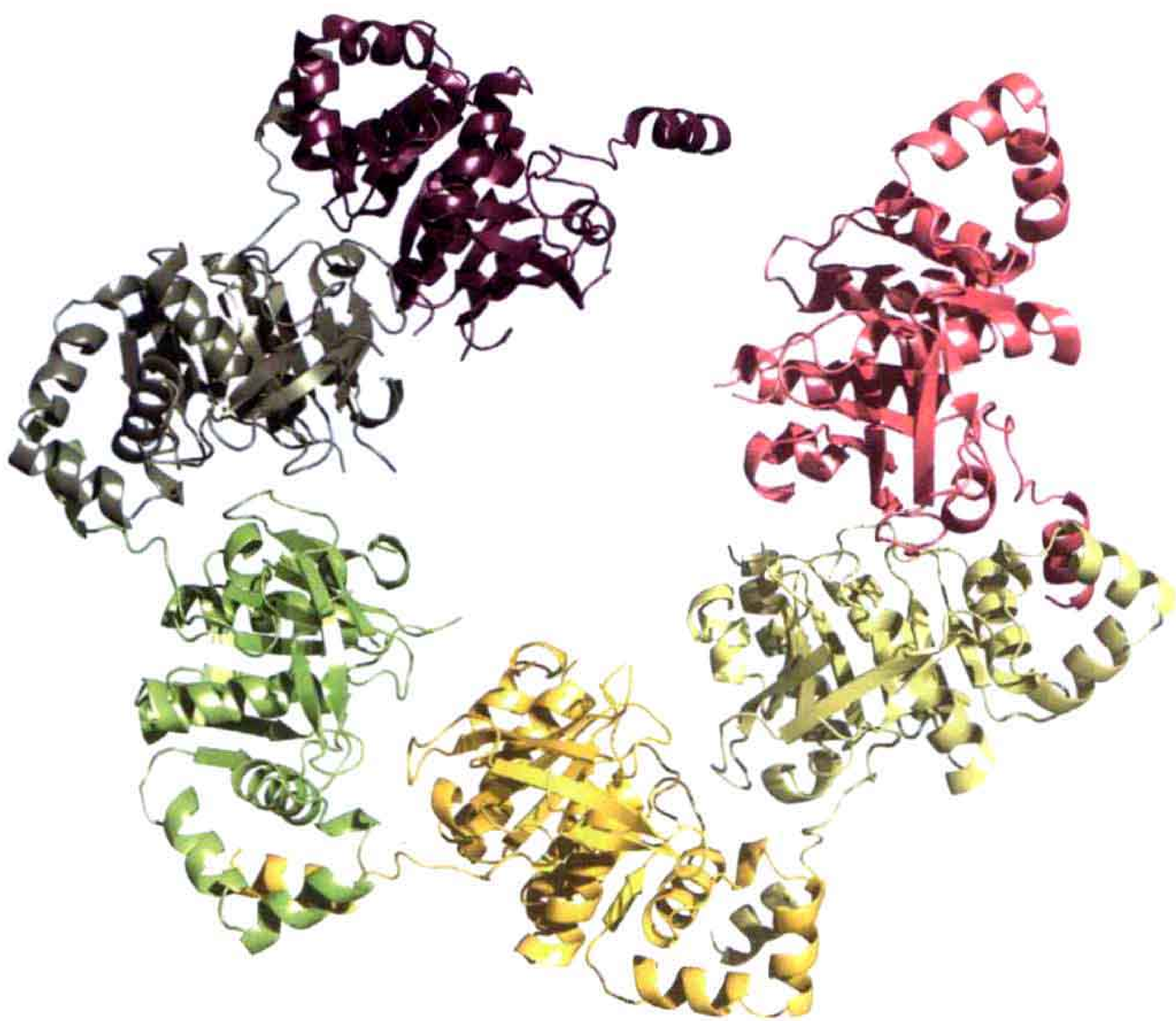


图 3.19　六聚溶菌酶的结构（PDB：1CRI）。酶的这一形式可以打开双链 DNA，为通过 DNA 聚合酶进行的复制或通过 RNA 聚合酶进行的转录提供可能。

3.3.1　核小体

DNA 和蛋白质形成的复合物在真核细胞中被称为染色质。结合到真核 DNA 上形成染色体的蛋白质可以被分为：①组蛋白染色体蛋白；②非组蛋白染色体蛋白，其中蛋白质占了染色质全部质量的一半。染色体组织的基本结构被称为核小体(nucleosome)。如果染色质被处理，导致局部展开后，可以在电子显微镜下观察到连续的小球在一条线上。每个小球代表了一个核小体，在每个小球之间的线被称为连接 DNA。每个核小体包含 DNA 的一个片段，通常含有大约 150 个碱基对，以及双份拷贝的 4 个组蛋白形成的蛋白质八聚体，其中这 4 种组蛋白分别为：H2A、H2B、H3 和 H4。双链 DNA 大约在八聚体蛋白上缠绕两圈（图 3.20 和图 3.21）。组蛋白含有非常丰富的精氨酸和赖氨酸残基，它们的正电荷抵消了来自 DNA 磷酸基团的负电荷。

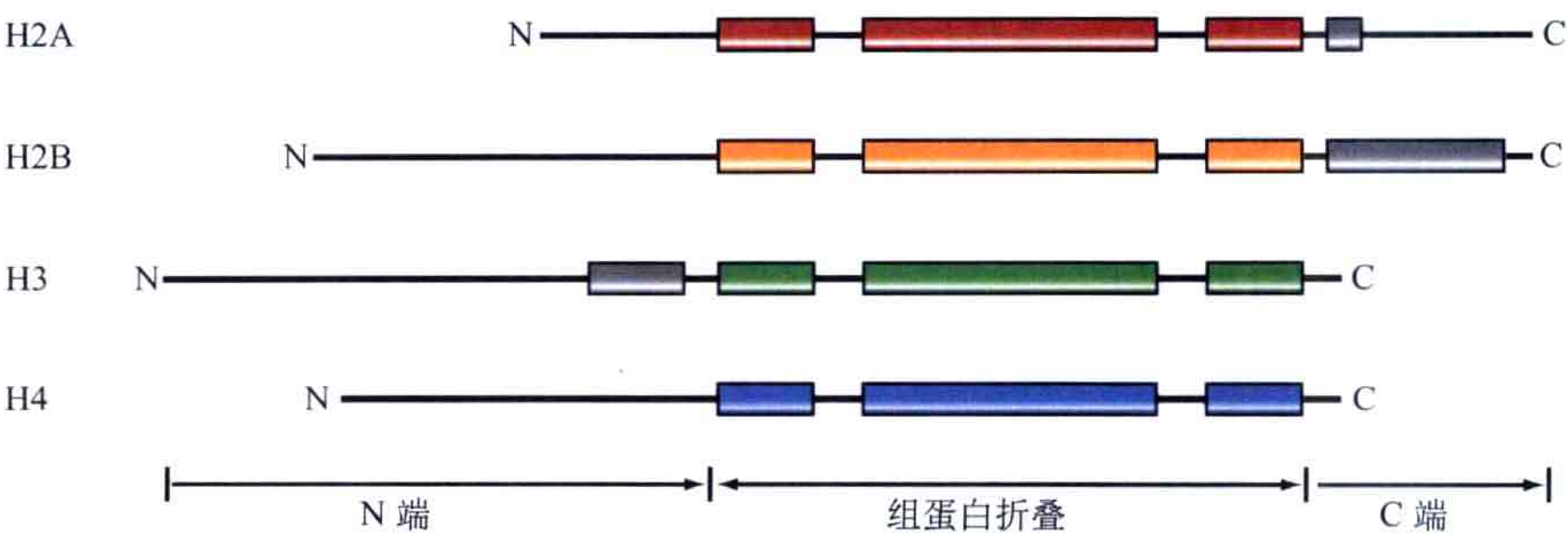


图 3.20　组蛋白各组成蛋白的示意图。它们都有一个组蛋白折叠中心区域和 N 端衍生链。

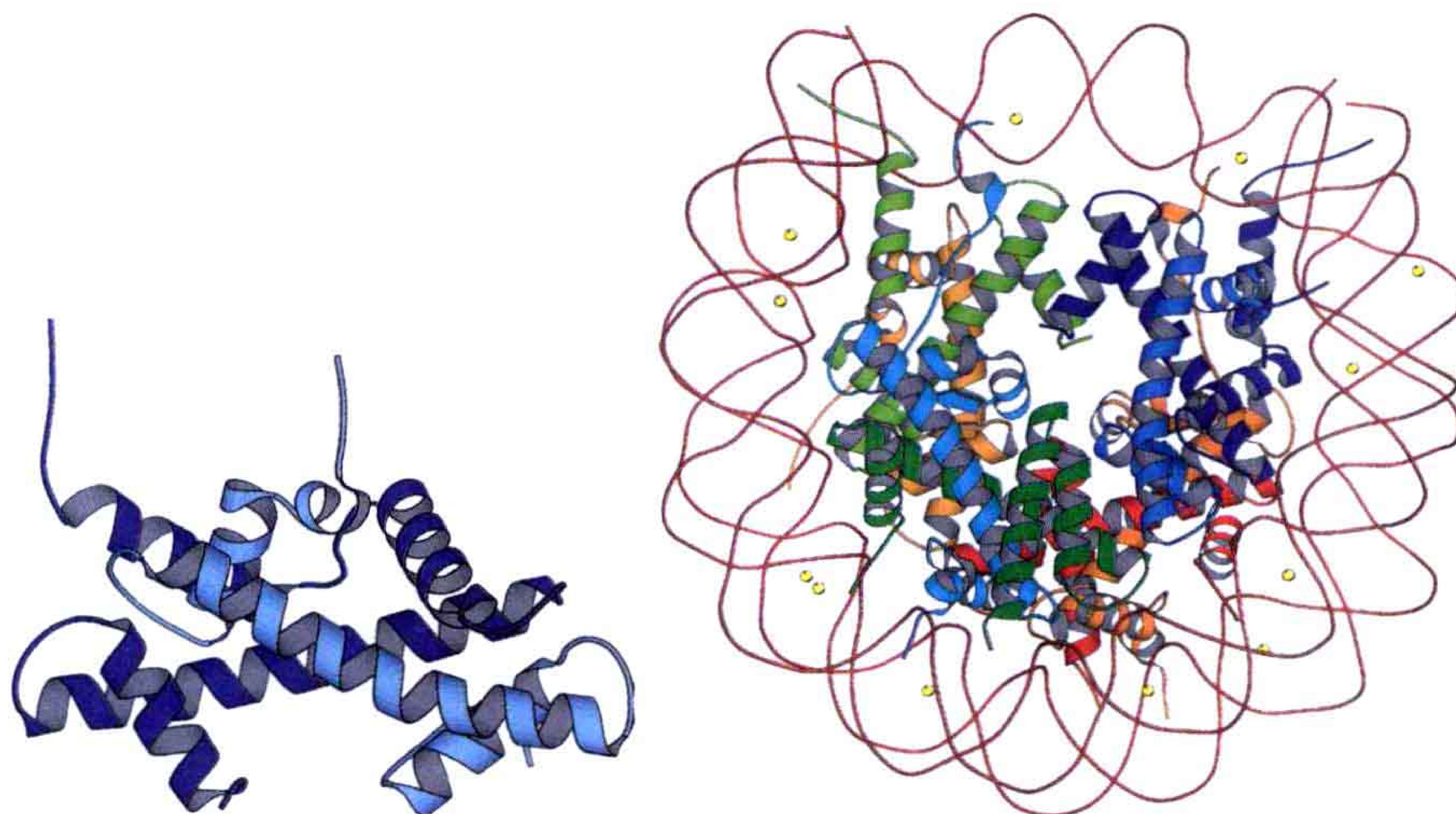


图 3.21 左：组蛋白 H3（蓝色）和 H4（亮蓝）组成的二聚体；右：核小体结构。由组蛋白组成的八聚体复合物中心和外部缠绕的 DNA 分子。中心位置的组蛋白亚基的颜色与图 3.20 中一一对应。

这 4 种组蛋白都是含有 102~135 个氨基酸残基的蛋白质，它们都含有一个相同的结构域，被称为组蛋白折叠（histone fold）（图 3.21）。这一折叠的中心位置可以用两段环形蛋白连接三条 α 螺旋来表示。这一结构在所有的真核生物中都是保守的。在组蛋白八聚体和相应的 DNA 螺旋之间存在数以百计的氢键、盐键和疏水相互作用。DNA 结构不是完全一致的，而是通过很多扭曲和畸形缠绕到组蛋白上的。

每个组蛋白核心都具有一个长 N 端尾巴，在细胞中会被修饰。这些改变显著地影响了核小体和它的高级结构，并且扮演了很重要的调控角色，如在基因表达过程中的调控。高度可变的组蛋白修饰可以影响 DNA 和组蛋白之间的相互作用，同时也存在于不同的核小体之间。除此之外，修饰的图形可以被其他蛋白质识别后结合。特别地，组蛋白修饰可以调控一个基因组的特定区域被转录（第七章和附录 E ）。组蛋白的修饰组合被称为组蛋白编码（histone code）。有人提出这些修饰是通过组合信号工作的，指导不同分子的结合和后续事件。当 DNA 被复制产生一个子细胞时，染色质的结构也将被拷贝，因此，组蛋白编码也将被拷贝。这一现象被称为表观遗传学（epigenetics）。

核小体会组装成更高级的染色质结构。第一层次的结构是 30nm 的纤维，外加的组蛋白 H1 在压缩结构中扮演了非常重要的角色。然而这个 30nm 的结构依然需要压缩至少为原来的 1/100 才能达到体内染色体结构的水平。但是对核小体的高级结构和全局结构依然知之甚少。

3.3.2 基因组结构

细菌通常只含有一条 DNA 分子，被称为基因组或染色体，它包含了细菌所有基因的信息。通常，成千上万的基因在基因组中可以被找到，但是依然只有很少一部分基因组是已知的，特别是在细菌需要宿主细胞来存活时。在真核细胞中，细胞核中的基

因会定位在一些不同的DNA分子，即染色体上。真核基因组需要携带至少5000个基因。这只是对于酵母来说的情况，但是在更加高等的生物体中，如昆虫、哺乳动物和植物，一个基因组中可能包含大于10 000个的基因。除此之外，真核细胞中的线粒体和叶绿体也含有DNA，但是这些分子要远小于细胞核中的DNA，并且只编码在这些细胞器中需要用到的蛋白质。

在20世纪80年代后期，研究者开始了对单细胞生物，如细菌和酵母基因组结构确定的研究。卓越的、里程碑式的工作是在1992年，第一个染色体的一级序列结构被测定成功。最小的酿酒酵母（baker yeast）染色体中的一条染色体——染色体Ⅲ的序列被Stephen Oliver带领的国际团体测定。在1995年，第一个完整的细菌基因组序列被测定，*S. cerevisiae* 的完整基因组序列在随后的一年中被测定。一些年之后，第一个动物果蝇（*Drosophila melanogaster*）基因组序列和第一个植物拟南芥(*Arabidopsis thaliana*)基因组序列被测定。很快，人类基因组测序也开始了。在2008年初，超过600个的细菌，超过50个的古细菌和超过70个的真核生物基因组序列已经被测定。全基因组测序的数量一直保持着迅猛的增长趋势。

被测定的基因组序列对于研究单个基因来说是非常有效的工具，同时这些序列也是从全局上去理解全基因组基因表达的可能的基础。基因组时代随着后基因组的发展变得更加重要，如转录组学、蛋白质组学等。从另一方面来说，由一级结构推测的核酸在活细胞中出现的更高级别的结构具有一定的代表性。尽管如此，依然需要做大量的工作，才能得到第一个染色体的更加细致的三维结构。

3.4 RNA的三级结构

3.4.1 RNA与DNA在结构和功能上的不同

虽然RNA在化学结构上和DNA非常相似，但是它们的三维结构却非常不同。DNA通常普遍地形成经典的Watson-Crick双螺旋结构，RNA却可以形成更大范围的结构折叠。一些RNA分子具有明确的、唯一的构型，可以很好地行使它们的功能，然而其他一些分子却可以形成更加柔韧的结构。其中tRNA分子和核糖体RNA分子（图3.22）就是RNA有序构型的例子。信使RNA的编码区域是一个柔韧性很大的RNA分子的例子，可以连续地穿过核糖体，在译码过程中局部二级结构需要被打开（或绕开）。具有双螺旋区域的分子利用各种相互作用类型来连接，单链区域通过连续的、远处的双螺旋片段相互作用。

从化学的角度讲，DNA和RNA之间的不同是不明显的：DNA由脱氧核糖核苷酸构建，而RNA由核糖核苷酸构建，它们之间的不同只是一个氧原子加到了核糖糖环的3′碳原子上。另一个不同是对碱基的选择不同：RNA采用尿嘧啶代替胸腺嘧啶，除了

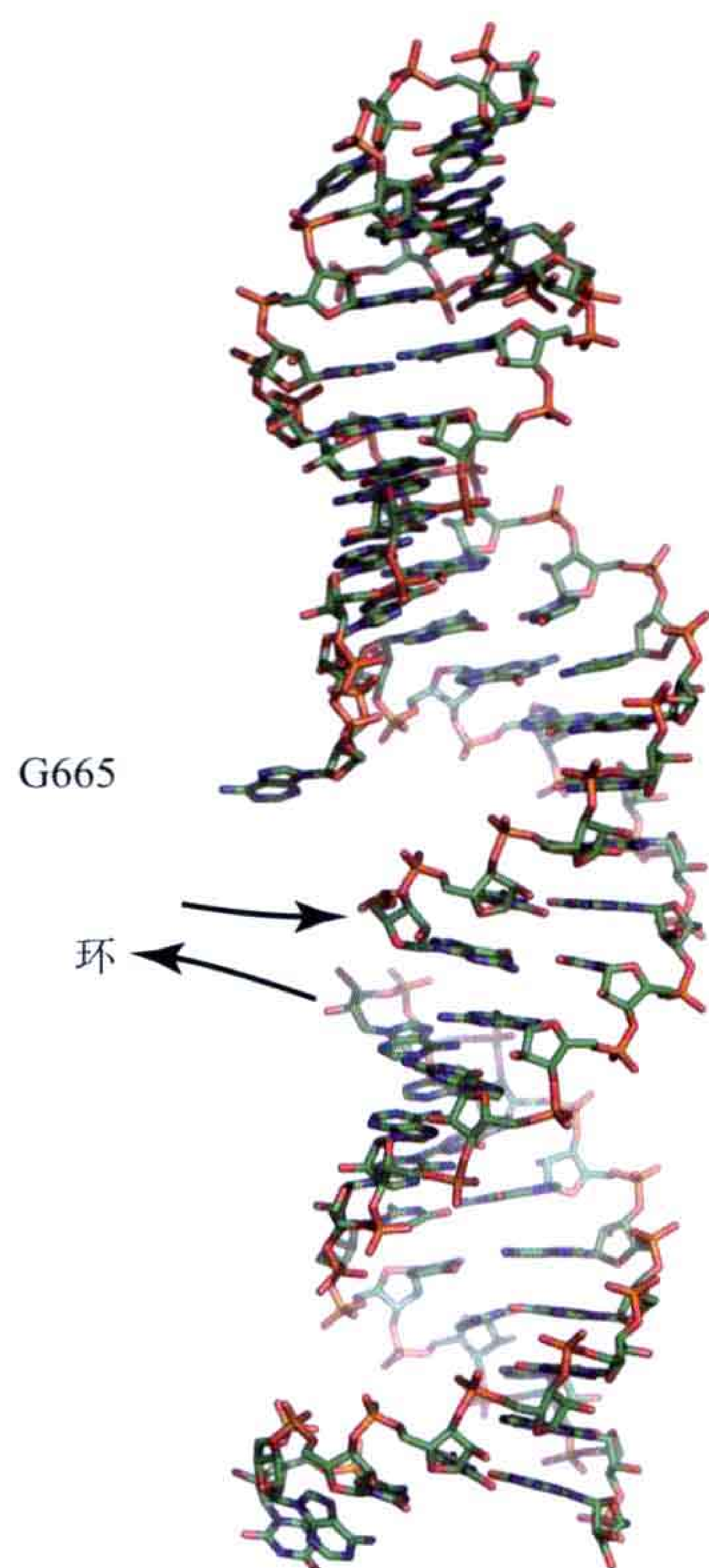


图 3.22 RNA（*T. thermophilus* 核糖体 16S rRNA 中的螺旋 23，PDB：1N32）。双链 RNA 一般接近 A- 型构象。

在胸腺嘧啶上多一个甲基基团外，其余都相同。但是这看上去很奇怪，为什么在 DNA 和 RNA 中不利用相同的 4 种碱基呢？这个问题需要从进化而不是仅仅从化学方面来回答。从尿嘧啶到胞嘧啶的化学降解是一种非常常见的 DNA 突变现象，但是这很容易被识别和修复。如果 DNA 利用尿嘧啶代替胸腺嘧啶，细胞将不会知道哪些尿嘧啶碱基需要修复。这在细胞把基因传递到下一代时，自我保护机制和 DNA 的保真度方面是非常重要的。

另一个相对于 DNA 的重要不同点是 RNA 没有互补链，经常以单链形式存在。尽管如此，RNA 经常含有伸展的自互补序列，可以回折叠后形成自身互补结构，形成一个伸展的发夹结构，并可以采取双螺旋结构。RNA 结构经常被看做包含了很多螺旋和线圈的复杂体系，一个发夹结构经常被称为“茎-环”结构（stem-loop）。

RNA 的螺旋区域经常具有 A 型结构，一圈含有 11 个碱基对。核糖具有 C3′-endo 折叠，以及大沟和小沟具有与 DNA 不同的形状。Watson-Crick 碱基对也能在 RNA 的双链区域找到，但是是另外一种形式的碱基配对或者碱基的排列。产生这一现象的原因是 RNA 分子没有包含完美的序列配对。通过碱基堆积形成双螺旋结构的趋势导致更少的氢键和更少的理想骨架构象来形成碱基对，如 GU 只有两个氢键，与正常的 Watson-Crick 碱基对配对原则有所差别。

在生物学上，RNA 具有很多明确的生物学功能：承载基因信息（mRNA，RNA 病毒中的病毒基因组）、形成结构实体（rRNA），具有识别任务（tRNA，siRNA）及催化化学反应（核糖酶）的作用。这些不同的角色起源于 RNA 分子可以采取很大范围内的三维结构的能力，这里重申了 RNA 比 DNA 更加具有柔韧性这一事实。但是为什么 RNA 会比 DNA 更加柔韧呢？其中一个原因可能是核糖基团具有一个额外的羟基基团，立体构象可能会更加被限制。从而，当脱氧核糖选择一个 2′-endo 糖折叠构象时，核糖会选择 3′-endo 糖折叠构象。这一结果在糖磷酸骨架的构象中更加线性，因此 RNA 可能更多地采取不同的构型。所以当核糖糖环完全刚性时，它依然具有更加柔韧的 RNA 骨架。

这一柔韧性的不同也同时在 RNA 双螺旋构型中得到了体现。这一改变了的糖折叠使磷酸基团连接的距离缩短了大概 1Å（图 3.23）。DNA 采取 B 型螺旋形式，一圈螺旋含有 10 个核苷酸；而 RNA 更倾向于 A 型，一圈螺旋含有 11~12 个核苷酸。在 DNA

中，碱基对的中心处于螺旋轴上。但在 RNA 双螺旋中，碱基对的中心偏离了螺旋轴大约 5Å。所有这些因素都导致了 RNA 双螺旋更加紧凑的堆积。

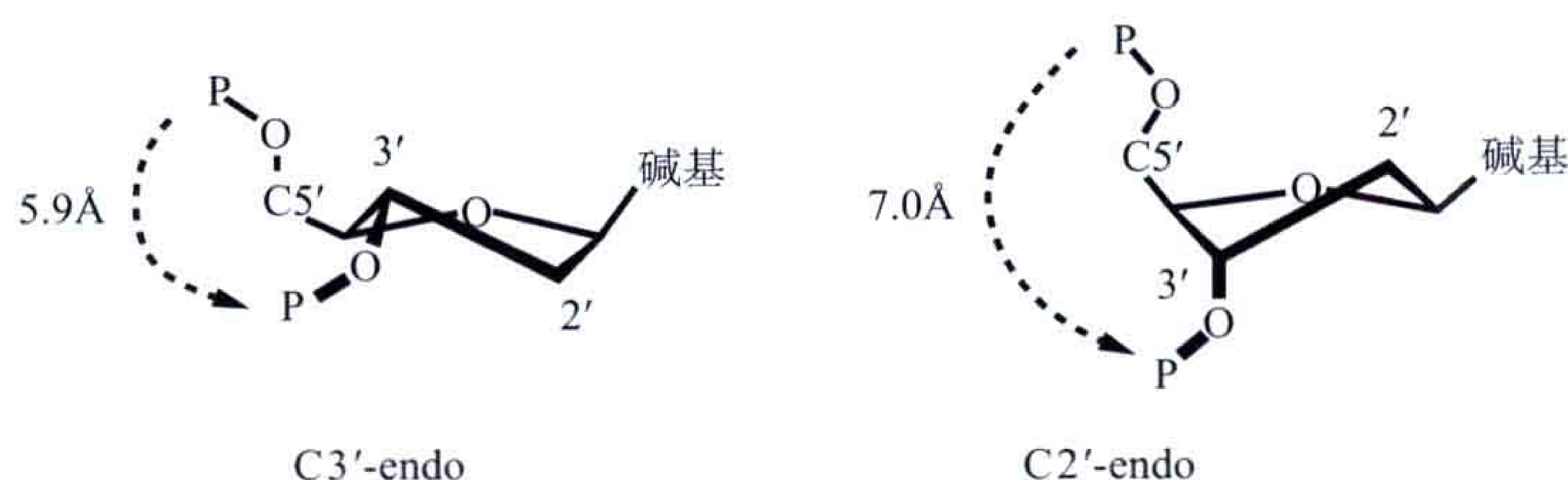


图 3.23　糖折叠在核酸中的两种最常见的形式。3′ -endo 折叠普遍存在于 RNA 和 A 型 DNA 中，而 2′ -endo 折叠是 B 型 DNA 的特征。可以看到，3′ -endo 折叠在骨架上具有更加短的磷酸 - 磷酸键，导致了更加紧凑的螺旋构型。

一个 RNA 螺旋的表面与 DNA 双螺旋的表面非常不同。RNA 的大沟更加窄和深，这是因为 RNA 大沟中的尿嘧啶残基没有胸腺嘧啶中的甲基基团。相反，小沟更加宽和浅。由于这个原因，大沟和小沟在 RNA 中更多地被称为深沟和浅沟。RNA 的深沟是阳离子、水分子和蛋白质侧链最容易结合的位点。

3.4.2　一级、二级、三级结构

我们可以像讨论蛋白质结构一样来讨论 RNA 不同层次的结构。它的二级结构由碱基配对决定。除了常规的 Watson-Crick 碱基对以外，不匹配的碱基对也能被找到，如 GU 碱基对。二级结构是由具有大约 −12kJ/mol 能量的氢键形成的。只是单个碱基对的构象倾向于能量不利。碱基对的堆积做出了大约 −23kJ/mol 的能量贡献，给出了最稳定的构象。碱基对的堆积能不是对称的，如 GC 在 AU 上的堆积与 AU 在 GC 上的堆积的能量贡献是不相同的。

最复杂的结构层次就是三级结构了。三级结构主要是二级结构单元在空间上的排列形式，如茎 - 环结构（stem-loop）。一个线圈可以和结构中另外一部分形成氢键，从而使其稳定。除此之外，范德华力和静电力也对三级结构具有很大的贡献，最后蛋白质分子可以与 RNA 三级结构相互作用并且使其稳定。一个基本的原则是，RNA 链像大部分的多肽链一样，不会形成拓扑结。

一个很大的典型 RNA 结构，其分辨率最多达到 2.5Å。这可能是因为实验上的困难，没法得到高纯度的原料来获得高质量的晶体。在中等层次的分辨率（2.5~3.5Å）下，磷酸和碱基平面可以被很好地定位，但是糖环，特别是它们的确切构象一般无法很好地在电子密度图中被确定。直到现在，对于 RNA 的结构只有很有限的信息可以被利用，这是因为在 RNA 结晶中遇到的实验困难。最近几年，随着核糖体晶体结构解析的成功，这一里程碑式的工作带来了 RNA 分子晶体结构高分辨率的一连串成就。

3.4.3 RNA 允许碱基配对的二选一

在 RNA 分子结构中关于碱基的一个有意义的内容是其存在通过非 Watson-Crick 碱基配对原则相互作用的碱基对。这些碱基经常涉及形成和稳定三级结构。这些二选一的碱基配对特定形式以 K.Hoogsteen 的名字命名，他于 1963 年第一个发现了这一现象。

虽然二选一碱基配对的方式有时候要比规则的 Watson-Crick 碱基配对原则具有更少的氢键，但是这增加了的堆积能量，补偿了这一能量的损失，从而在能量上变得有利。RNA 就是这样可以适应很多不同的结构需求，同时还能保持理想的稳定性。含有二选一碱基对的位点通常是生物学上重要的位点，因为这些不完美的氢键可以参与与蛋白质或其他因子的相互作用。

Leontis 和 Westhof 提出了一个图表，试图使非规范碱基配对更加合理化，此图表中嘌呤和嘧啶参与碱基配对的侧面（edge）都被确定。嘌呤有三个侧面 [Watson-Crick (WC)、Hoogsteen (H) 和糖 (S) 侧面]，而嘧啶只有两个（图 3.24）。一个碱基对可根据其所参与碱基配对的侧面来分类。一种标准的碱基对就是 WC/WC，而 Hoogsteen 碱基对可以是 H/WC 或者 WC/H。如果组合嘌呤上的三种侧面和嘧啶上的两种侧面，就有 6 种组合方式出现，再加上如果平行和反平行链的方向被考虑，则全部的碱基对都能被分类到 12 种可能的碱基配对家族中（表 3.1）。如果糖苷键在通过碱基对平面的对称线的同一侧，则这个碱基对被称为 *cis*；相反，这个碱基被称为 *trans*。

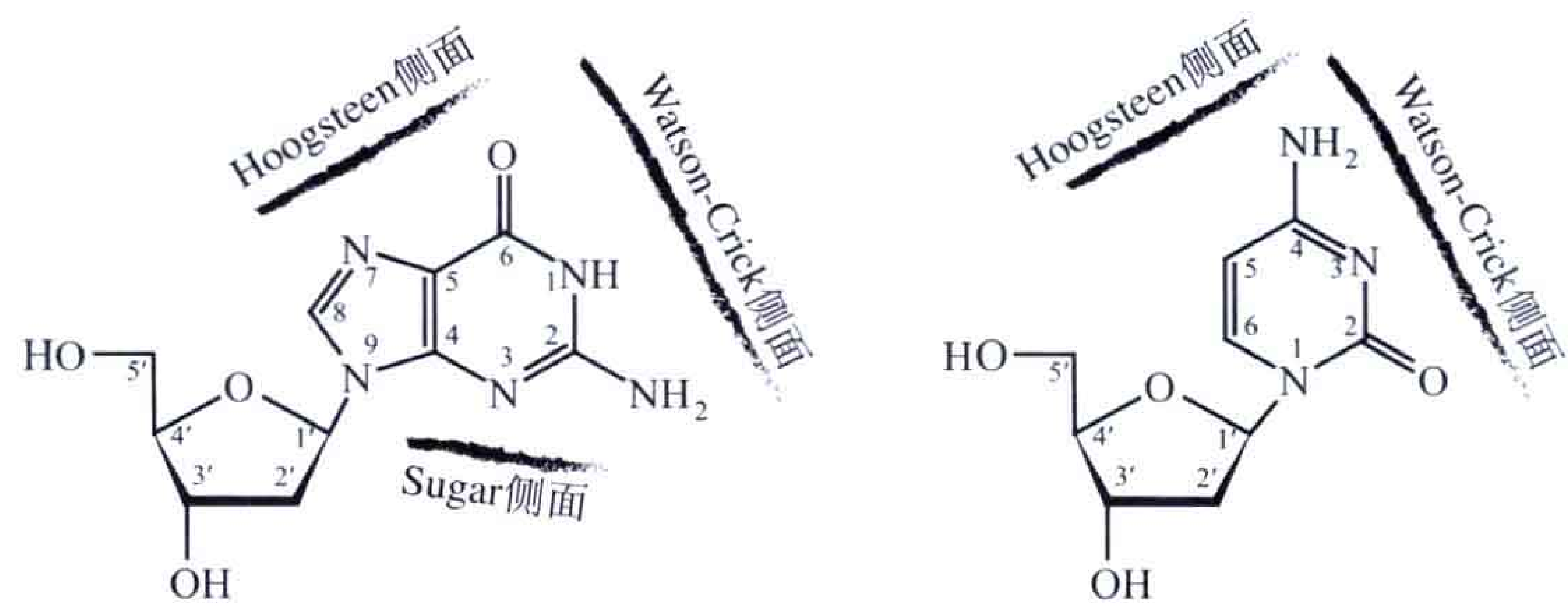


图 3.24 碱基对中嘌呤和嘧啶中侧面的定义。

表 3.1 Leontis/Westhof 碱基对分类中碱基配对的 12 个主要家族

序号	糖苷键的方向	参与相互作用的边缘	链之间的方向	符号
1	*cis*	Watson–Crick/Watson–Crick	反平行	
2	*trans*	Watson–Crick/Watson–Crick	平行	
3	*cis*	Watson–Crick/Hoogsteen	平行	
4	*trans*	Watson–Crick/Hoogsteen	反平行	
5	*cis*	Watson–Crick/Sugar Edge	反平行	
6	*trans*	Watson–Crick/Sugar Edge	平行	
7	*cis*	Hoogsteen/Hoogsteen	反平行	

续表

序号	糖苷键的方向	参与相互作用的边缘	链之间的方向	符号
8	*trans*	Hoogsteen/Hoogsteen	平行	
9	*cis*	Hoogsteen/Sugar Edge	平行	
10	*trans*	Hoogsteen/Sugar Edge	反平行	
11	*cis*	Hoogsteen/Sugar Edge	反平行	
12	*trans*	Hoogsteen/Sugar Edge	平行	

注：符号在 *cis* 构象时被完全填充，而在 *trans* 构象时是中空的。圆圈表示 Watson-Crick；方框表示 Hoogsteen；三角形表示糖边缘。

在经典的定义中，Hoogsteen 碱基配对一般经常包含一个与嘌呤 Hoogsteen 侧面的相互作用，在鸟嘌呤上含有两个氢键受体（N7 和 O6），在腺嘌呤上含有一个氢键受体和一个氢键给体（N7 和 N6）。在相反的 Hoogsteen 碱基配对中，这两条链是平行的，意思是嘧啶配对是倒装的，而糖环结束在 *trans* 的位置上。

图 3.25 表示了一个标准的 Watson-Crick GC 碱基对（*cis* WC/WC）相对于 GC 反碱基对（*trans* WC/WC），可以在平行的骨架链上被观察到。因为 RNA 平行链没有形成常见的茎环折叠结构，反 Watson-Crick 相互作用是三级相互作用。

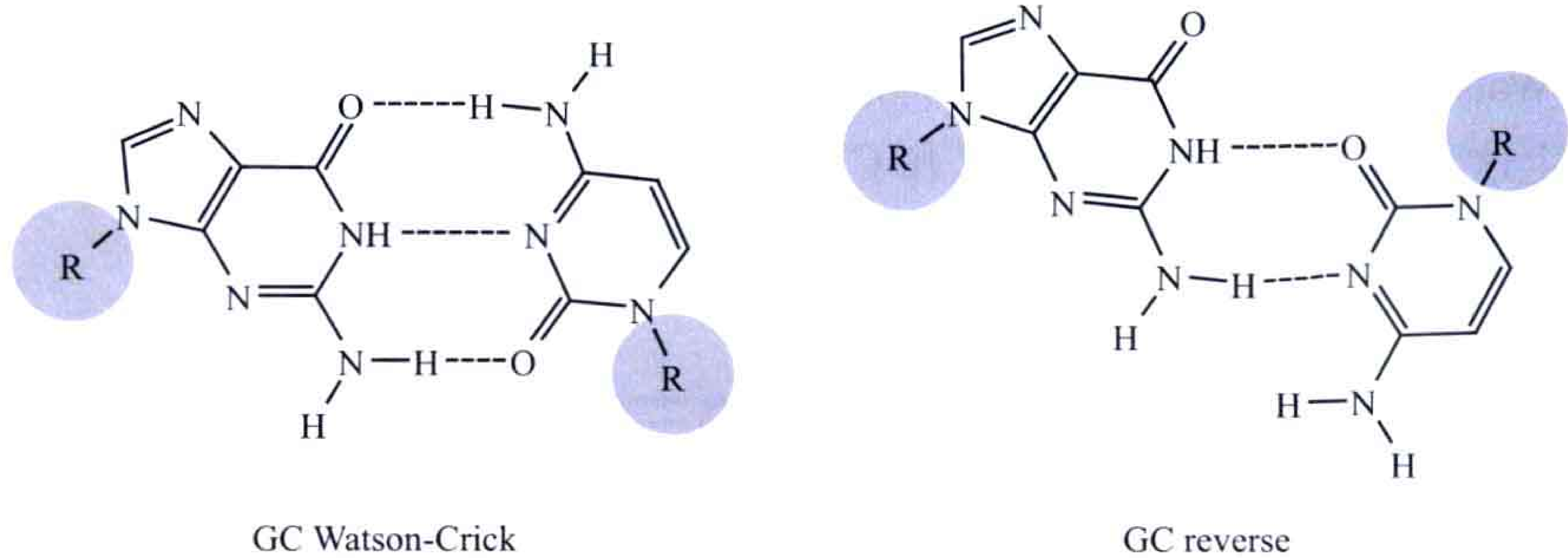


图 3.25　左：规则的 Watson-Crick GC 碱基对（*cis*）；右：GC 反 Watson-Crick 碱基对（*trans*）。

另一个事实是，从 RNA 的三维结构上看，骨架之间的距离在 *cis* 和 *trans* 构型中是不同的。虽然没有在可测量的范围内，但这样的例子还是可从图 3.26 中的图解中看到。骨架链之间的距离在 AU Hoogsteen 碱基对中要比 AU 反 Hoogsteen 对中短。

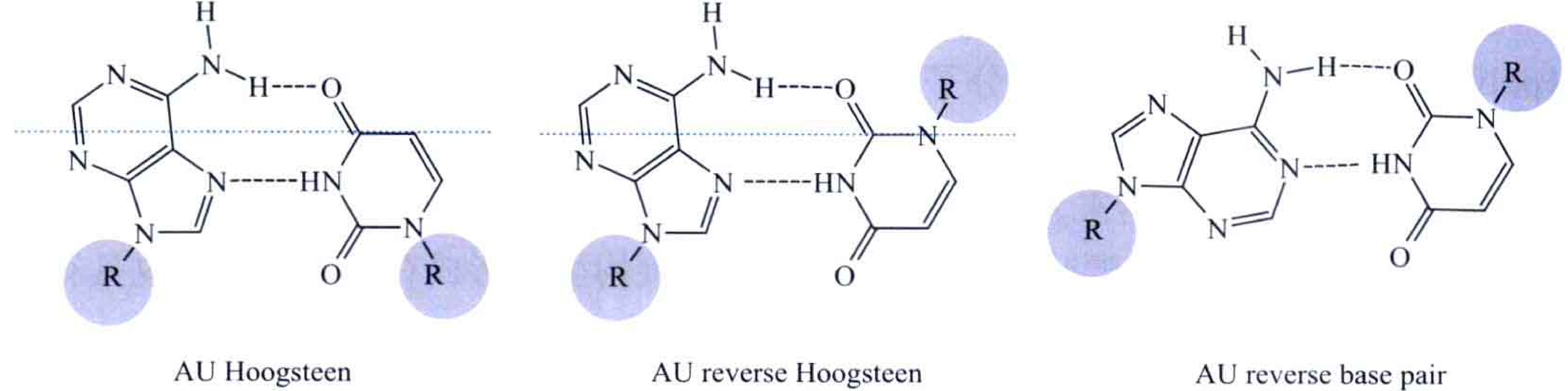


图 3.26　左：AU Hoogsteen 碱基对；中：AU 反 Hoogsteen；右：AU 反碱基对。蓝色虚线表示对称线，用来定义碱基对的 *cis*/*trans* 构型。AU Hoogsteen 碱基对就是 *cis*-H/WC，AU 反 Hoogsteen 就是 *trans* H/WC。

Francis Crick 提出了一个术语，即“wobble”（摇摆）碱基对，用来表示在密码子和反密码子相互作用中非互补的 GU 碱基配对，这可以由基因编码的简并来加以说明。图 3.27 描述了 GU wobble 碱基对，这是最常见的二选一碱基配对模式；而 GU 反 wobble 中尿嘧啶基团绕着胺氢键轴做了简单反转。GU′ wobble 碱基配对导致鸟嘌呤氢键的丢失，但是空闲的胺基团经常能与周围其他的碱基形成氢键，这里假设其与周围的亚胺基团取向相一致。GU wobble 碱基配对可以被看成是规则的 Watson-Crick 模式，只是被替换了嘧啶成员而已。

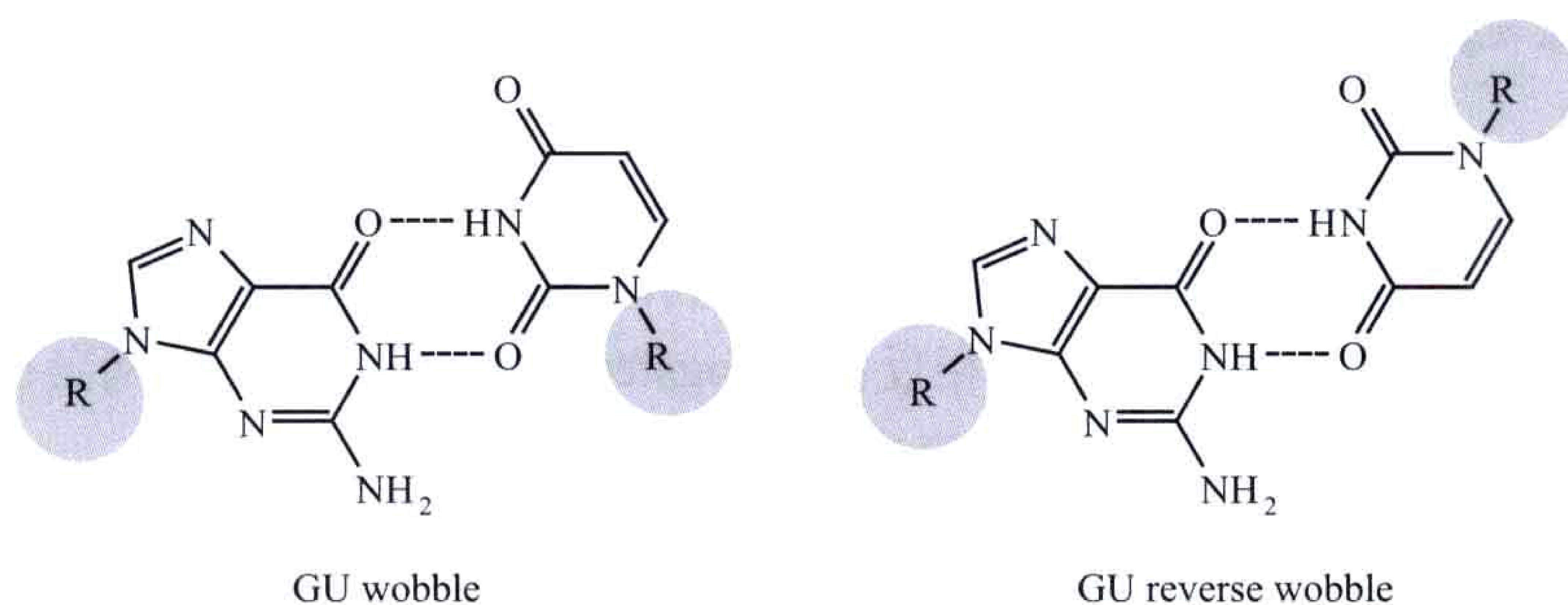


图 3.27　左：GU wobble；右 :GU 反 wobble。

GU wobble 配对的几何性质可以使它形成非常整齐的 A 型螺旋，因此其经常会取代整齐的 Watson-Crick 配对形式。在特定的化学环境中，A 碱基的 N1 位置可以被质子化。当发生这种情况时，A 碱基可以和 C 碱基形成氢键，形成一个（A+）：C 碱基对。（A+）：C 碱基对在几何上是和 GU wobble 碱基对空间异构的。

3.4.4　碱基三联码：一个突出的三级结构序列

在前面章节中描述过的二选一碱基配对模式提供了丰富的多碱基相互作用选择性。当一个碱基对由两个核苷酸形成时，一些氢键给体或受体依然可以被利用，因此会出现与蛋白质氨基酸的侧链或其他的核苷酸的交替相互作用，这些多碱基相互作用中最重要的就是碱基三联码，其在维持 RNA 分子的三级结构中也很重要，一些三联碱基相互作用的例子将在下一章中描述。

三联码相互作用的可能性在 1957 年第一次被 Felsenfeld 实现，Felsenfeld 证明一个（poly-A）：（poly-U）双倍体分子可以与第二个 poly-U 链相互作用，形成一个三链复合物（图 3.28）。增加的 poly-U 链与二倍体的大沟通过与 poly-A 形成 Hoogsteen 碱基对而进行相互作用。后来发现一些其他的序列组合也可以形成三聚螺旋构型，只要它们含有两个嘧啶和一个嘌呤，如 C：G-C。三螺旋构型的形成在基因修复中也具有很重要的意义。

另一个高度有序的结构是四聚复合物。4 个尿嘌呤残基可以连接在一起形成 G-四聚体（G-tetrad），它是平面循环的氢键结构（图 3.29）。如果一个核酸序列具有连续的两个或更多的鸟嘌呤碱基，则 4 个分开的链可以联合到一起，形成一个平行

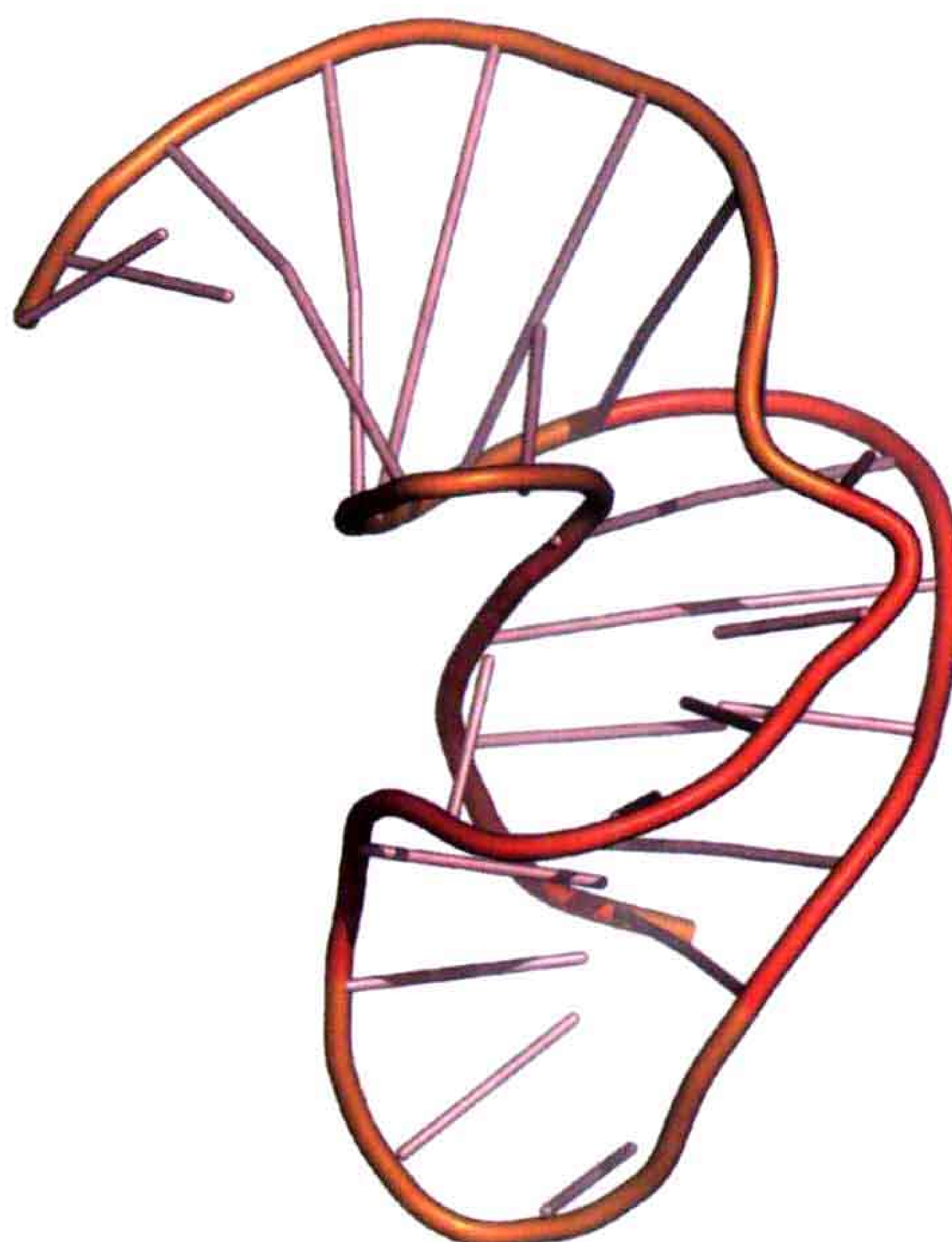


图 3.28 一个框架移位的假结中的三聚 RNA 螺旋。

的四聚体的四重复合物（quadruplex 或者 tetraplex）。四重复合物的另一个形式在两个 G-发夹联合到一起时形成，最后，如果一个序列含有 4 个分开在 4 个地方的鸟嘌呤，就能形成一个内在分子反平行四重复合物。这些不同的四重复合物形式如图 3.30 表示，其中一个在图 3.31 中有更为详细的描述，四重复合物显示了一个不同寻常的基于特定金属离子的构象和稳定性。可以诱导形成四重复合物结构的阳离子有很多，包括 NH_4^+、Tl^+、Sr^{2+}、Ba^{2+} 和 Pb^{2+}。

DNA 序列在染色体中可能形成众多的四重复合物（蛋白质 -DNA 复合物，给线性染色体的末端加帽，6.1.3 节）。这里有很多由末端着丝粒蛋白捆绑成四重复合物的例子。另一个显示具有与 DNA 四重复合物相互作用的蛋白质是人类 DNA 拓扑异构酶 I。这种酶可以结合到四链四重复合物上，也可以结合到单一分子的四重复合物上，除此之外，也可以诱导四链四重复合物结构的形成。

四重复合物形成序列也在免疫球蛋白开关区域和基因启动区域被找到。另外，四重复合物也在胰岛素样生长因子Ⅱ（insulin-like growth factor Ⅱ）的 mRNA 中被识别到，这被认为是维持侧面分裂位点的，或者可能是抑制和调节附近基因转译的。

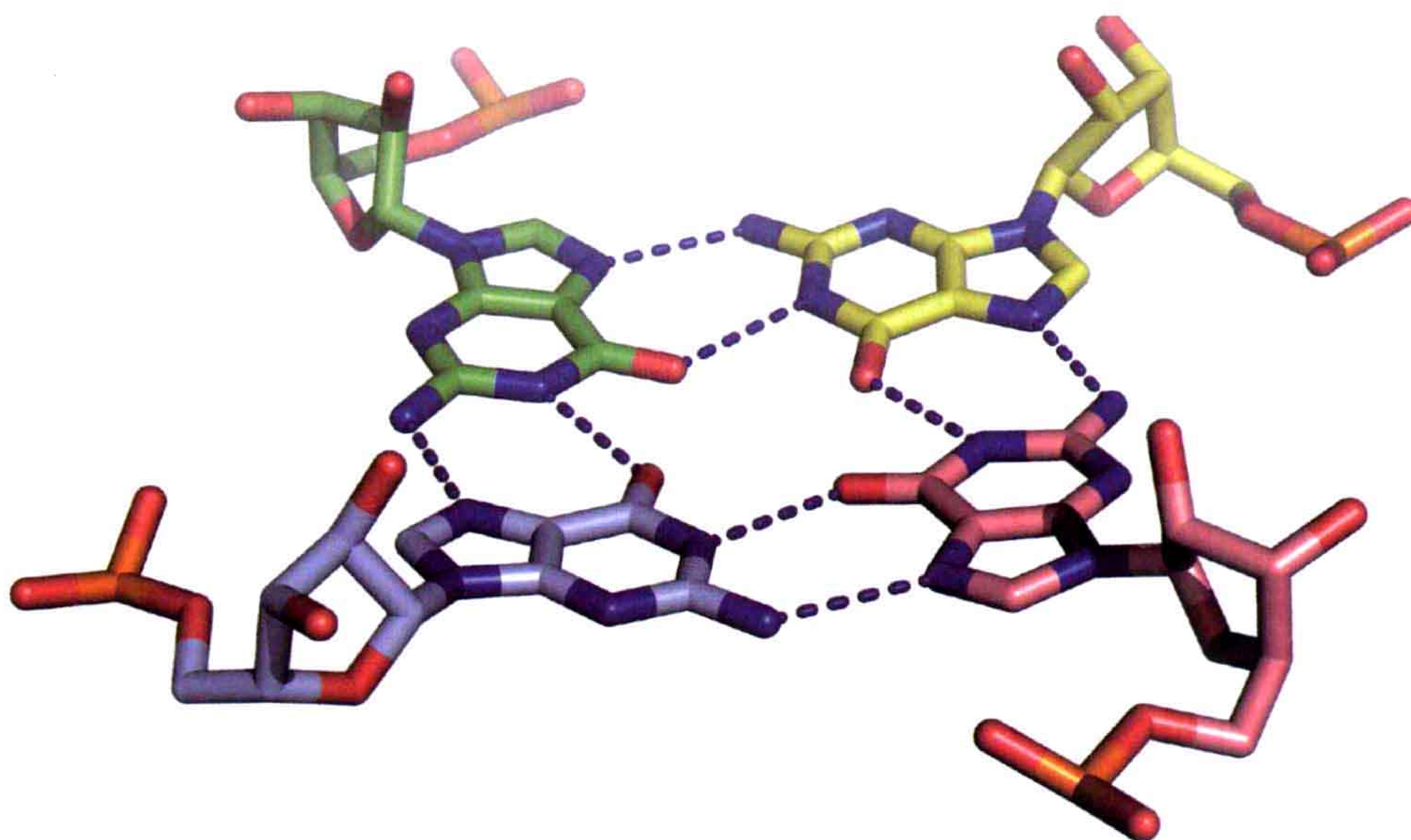


图 3.29 4 个鸟嘌呤残基形成的 G- 四聚体。

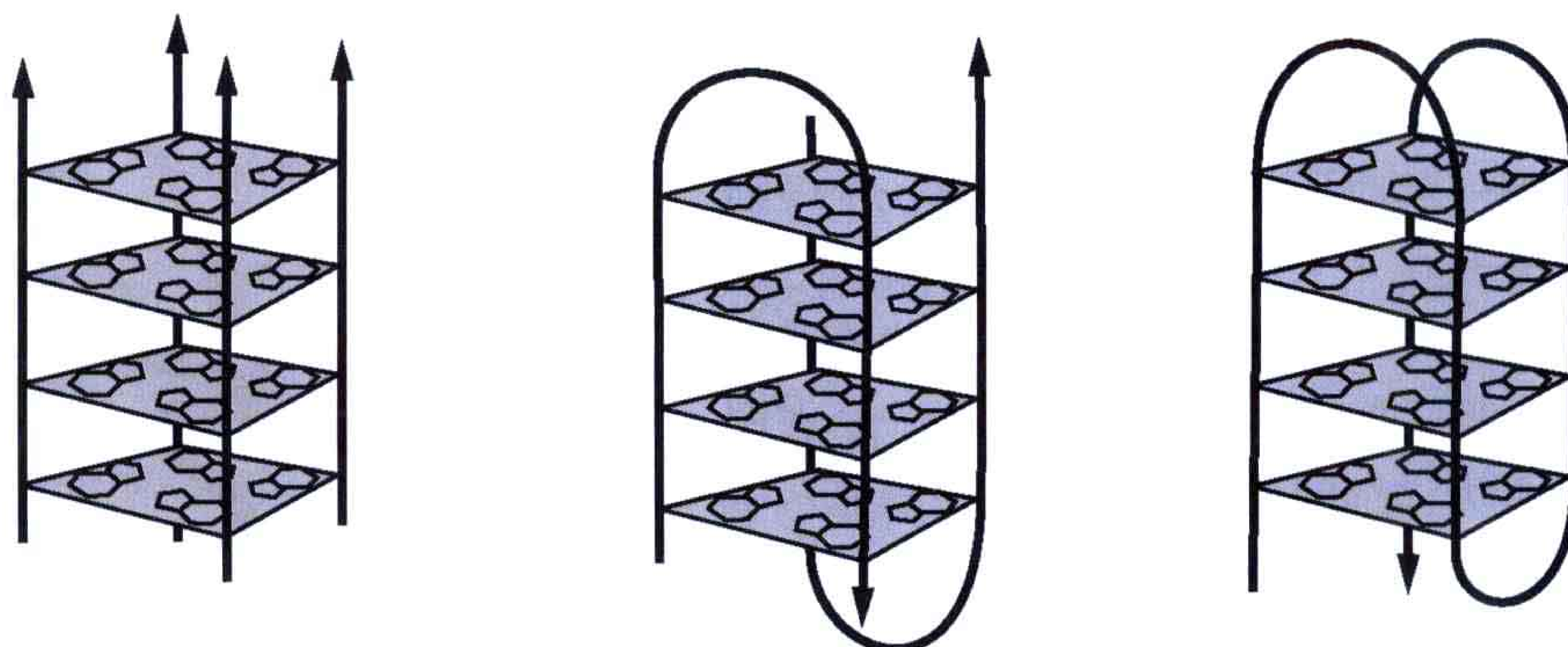


图 3.30　三个基本的四重复合物结构。左：平行四聚体四重复合物；中：两个 G- 发夹组成的四重复合物的一些形式中的一种；右：分子内反平行四重复合物。

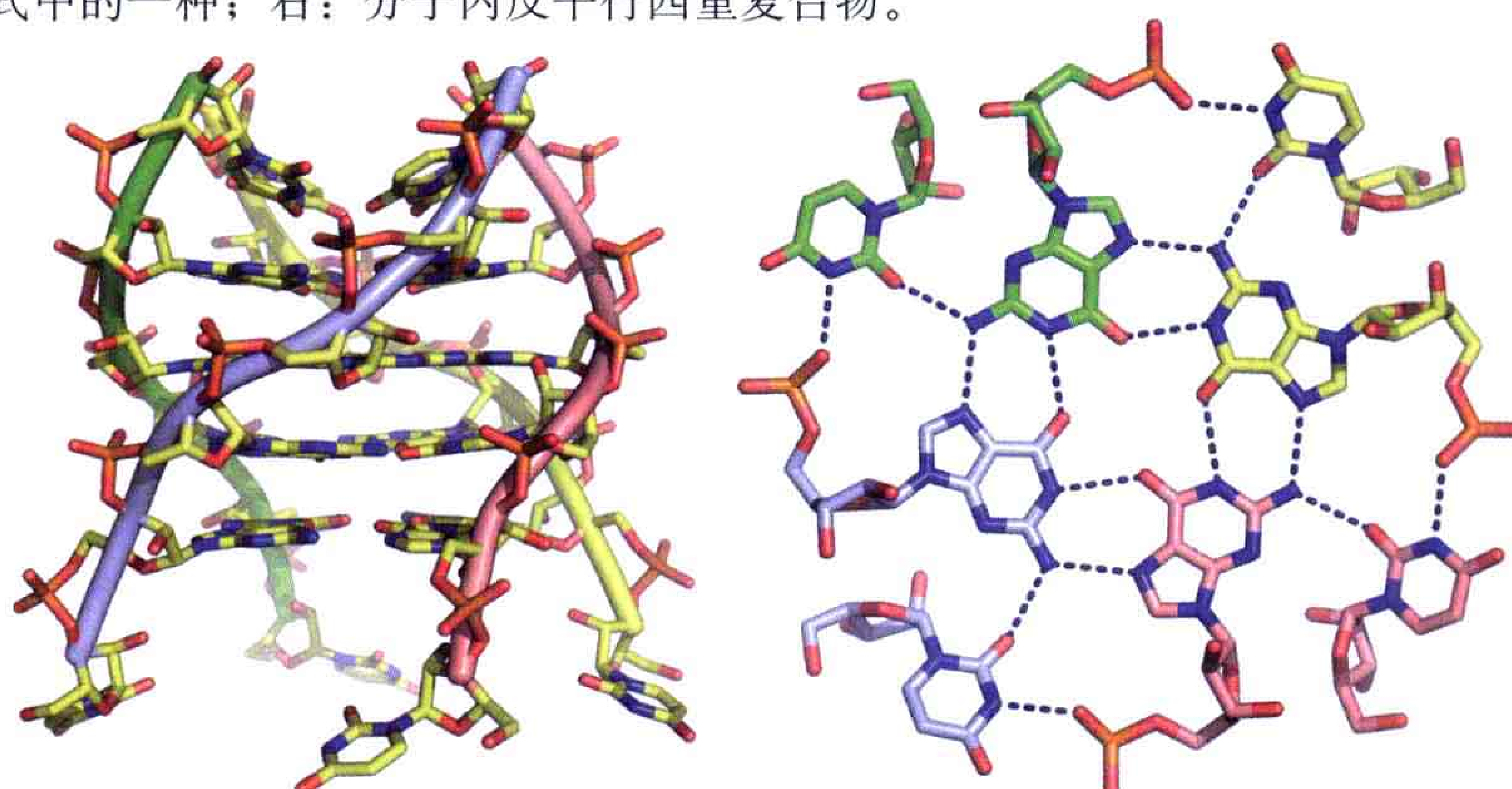


图 3.31　左：RNA 四重复合物；右：来自同一结构的鸟嘌呤和尿嘧啶碱基的八聚体（PDB：1J8G）。

3.4.5　RNA 含有的修饰碱基

RNA 的转录后修饰使其中的修饰碱基比 DNA 中的更加多样化。这在功能性 RNA 中尤其显著，如 tRNA 和 rRNA，它们的核糖和碱基都能被修饰。这些碱基可以彻底地改变 RNA 分子的化学性质，可以对分子的稳定性作出贡献，同时参与它的外部相互作用和化学反应。

关于转录后修饰重要性的一个很好的例子是大肠杆菌异亮氨酸特定 $tRNA_2^{Ile}$，特定编码为 AUA。tRNA 包含了修饰的碱基，即赖西丁（lysidine），在反密码子的第一位上。这种碱基是一种胞嘧啶，它是在转录后修饰中在 C2 位置加了一个赖氨酸生成的（图 3.32）。如果这个赖西丁残基被一个天然的胞嘧啶取代，异亮氨酸将会显著减少，而出人意料地变成了甲硫氨酸被更多地接受。这是怎么出现的呢？在大肠杆菌中原来是这样的，如果赖西丁残基在反密码子环区出现，同族的异亮氨酸 tRNA 连接酶只会识别和记录 $tRNA_2^{Ile}$，而且，这个 $tRNA_2^{Ile}$ 的反密码子序列是 CAT，这是甲硫氨酸的标准编码。赖西丁的缺失，将导致 tRNA 被 CAT 识别的甲硫氨酸基 tRNA 连接酶识别和错记。因此单个的转录后修饰是这个 tRNA 的编码和氨基酸特异性共同的原因。

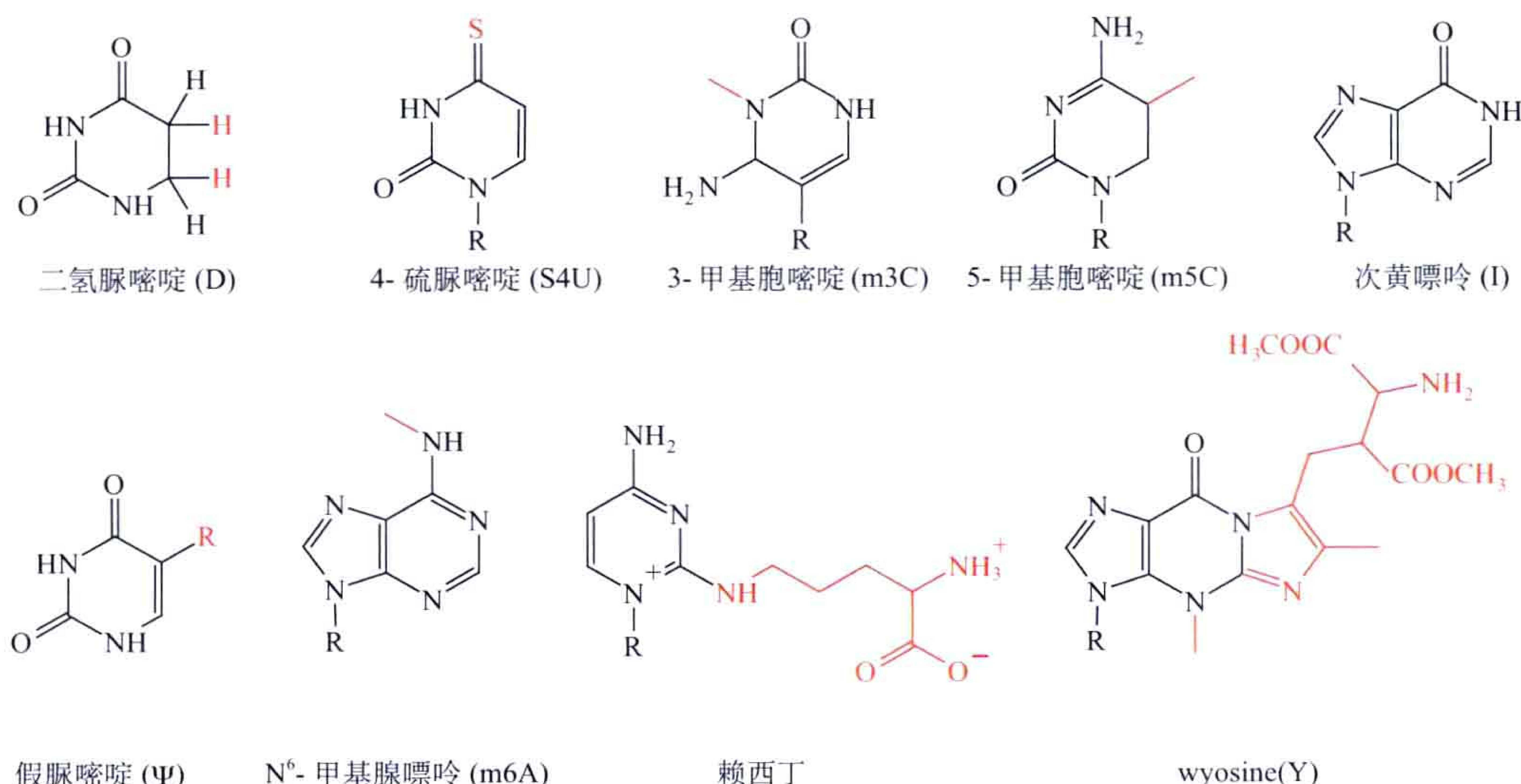


图 3.32　RNA 中修饰碱基的例子。修饰部分被标记为红色。R 表示核糖。

3.4.6　互补碱基突变揭示的 RNA 二级结构

来自不同物种的、同样的 RNA 分子的一级结构就像在蛋白质中的情况，可以有很大的变化。尽管如此，它们的结构却非常少。在 RNA 结构中主要的部分是序列中可以伸展成对形成螺旋的部分（图 3.33）。通常，碱基配对问题存在可选择的解决方案，其具有形成不同的和冲突的螺旋结构排列的风险。双螺旋结构中出现保守序列的频率往往比单链区域少，因为单链区域经常具有不同的功能。这在 tRNA 和核糖体 RNA 中已经被阐述。在 RNA 双螺旋的预测中这是一个主要的问题。然而，此保守结构要求碱基配对具有保守性从而形成双螺旋构型。在一条链上的改变，将导致另一条链上一致的改变。因此二级结构的预测需要多相关序列的信息。

RNA 的二级结构可以从大量生物体的许多相关 RNA 序列比对中被精确地确定。最根本的假设是当突变在 RNA 中出现时，RNA 还会保持二级和三级结构。如果一个点突变出现在双螺旋区域，螺旋将失去一对碱基，因此会失去稳定。这样的突变一般由相反链上的另一个碱基改变来补偿，因此螺旋碱基配对依然被保留。所以，在序列比对中寻找互补的碱基改变，对于寻找可能形成双螺旋不变结构的区域是可行的办法。

当这些包含有 Watson-Crick 碱基配对作用时，RNA 的二级结构一旦被推测，相同的方法可以被用来探测三级结构中的相互作用。然而，当 RNA 中三级结构的相互作用是非规则时（而且经常出现），或者如果相互作用中包含一个序列的伸缩，而其他条件都一样时，这样的共同变种是没有可能得到的。因此，对于 RNA 三级结构中的相互作用是无法确定的，而且这还是一个理论上的挑战。

图 3.33　大 RNA 的二级结构非常复杂。图中所示为大肠杆菌（*E.coli*）中 1542 个核苷酸长的 16S rRNA 结构。

3.4.7　RNA 结构模体（motif）

就像蛋白质分子的所有三维结构都是由各种各样的 α 螺旋和 β 折叠的组合组成的一样，RNA 的二级结构模体（motif）提供了它的结构骨架。这些模体主要有两个类别，即发夹结构（hairpin loop）和内环结构（internal loop）。发夹环是在 RNA 序列的可伸展区域是一个完全自互补的序列时形成的。例如，在 23S rRNA 中被称为 A 环（A loop）的一段序列（见 8.5 节）



最前面的 7 个碱基与最后的 7 个碱基是完全互补的，因此这样的序列可以形成一个发夹环结构。这样的发夹环，或者称茎环结构（stem-loop structure），是根据未配对区的碱基数来分类的，被分别称为三环、四环、五环（triloop、tetraloop、pentaloop）等（图 3.34）。其中一个因素对于形成发夹环结构具有非常突出的贡献，就是堆积力。在上面的例子中，可以想象这个结构可以被假设成一个五环的结构，但是事实上堆积力会促进双螺旋中形成连续的、可选择的 CU 碱基配对，这已经在 NMR 的研究中得到了证实，核糖体中 A 环的确是三环结构的。

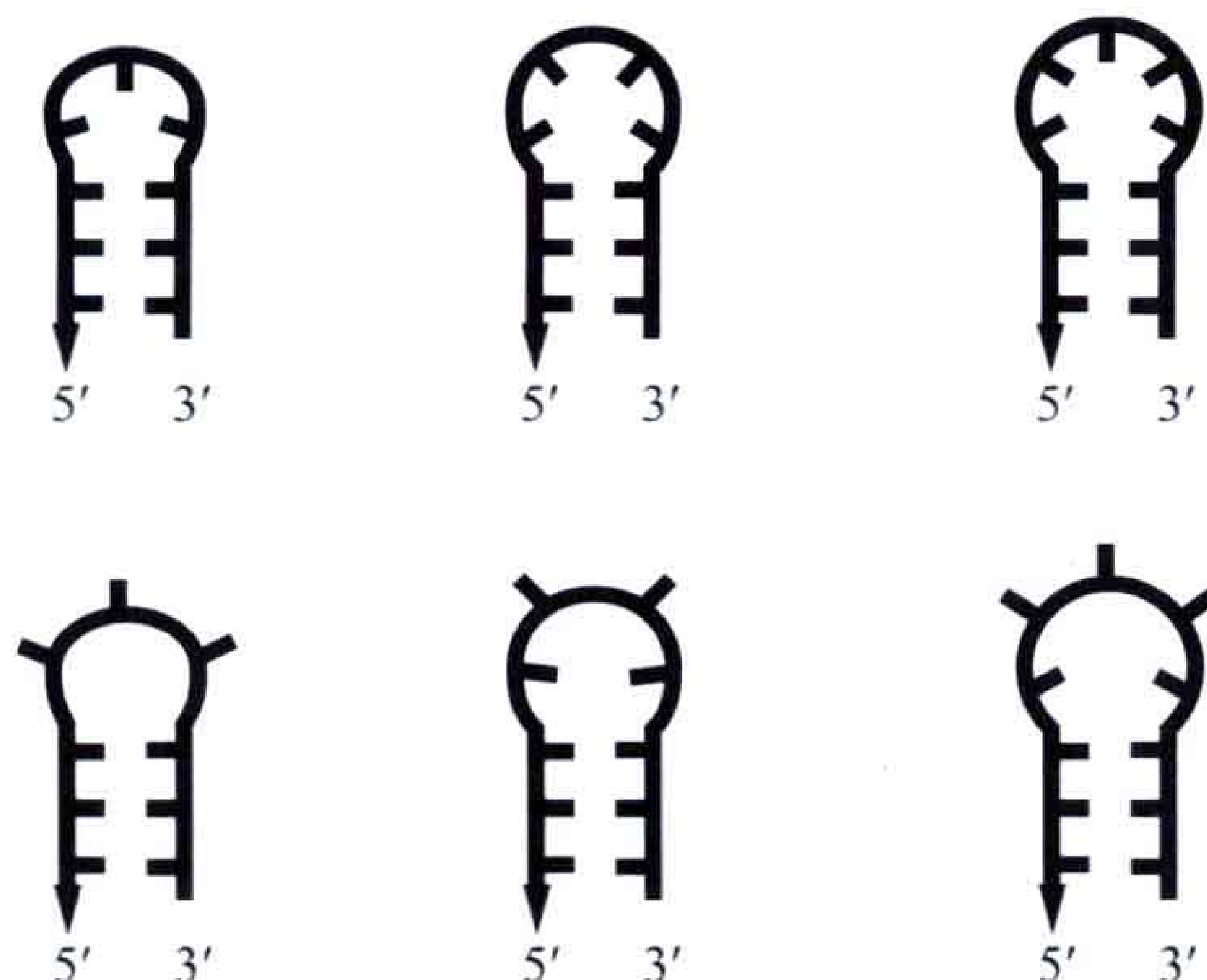


图 3.34　不同的发夹环结构。上行，从左开始：在未配对环中含有 3 个（原文为 two）碱基的三环结构（triloop）、4 个碱基的四环结构（tetraloop），以及 5 个（原文为 four）碱基的五环结构（pentaloop）；下行，从左开始：和上行同样的结构，只是一部分碱基在外侧——三环、四环（两个在内，两个在外）、五环结构（两个在外，三个在内）。



图 3.35　接吻环（“kissing loop”）的表述图。

未成对的碱基倾向于形成双螺旋的堆积，因此在环区未成对的许多碱基后面，发夹环可以根据未成对碱基相对于双螺旋如何排列来分类，如碱基是环外的（looped-out）还是堆内的（stacked-in）。例如，三环的一些分类，根据三个、两个、一个或是零个碱基在主要的堆积区内。此外，如果出现堆积内碱基，那么经常会在 5′ 或 3′ 端堆积区。

在特殊情况下，两个发夹环包含互补碱基序列，它们可以形成三级结构的相互作用，很形象地被称为“接吻环”（图 3.35）。这一结构模体在 23S rRNA 中被发现，也是 HIV-1 基因组中的一个结构组件。

另一个关于接吻环的例子出现在 $tRNA^{Asp}$ 的晶体结构中，其中含有反密码子 GUC。这里有两个 tRNA 分子相互作用通过在它们中间位置含有一个 UU 错配的二倍体的反密码子环区。这一二倍体被证明一直处在不断变化中。例如，在酵母 $tRNA^{Asp}$（GUC）和大肠杆菌 $tRNA^{Val}$（GAC）复合物之间，其中 tRNA 含有互补的反密码子。

RNA 的骨架中所有残基含有 6 个自由度，而多聚肽骨架只含有 2 个自由度。这种非常柔性的结构允许单链 RNA 区域采取非常多的构象。然而，一部分单链 RNA 模体被认为经常出现。首先是 S- 转向（S-turn），它是磷酸糖骨架形成了两个连续转弯的 S

形结构，被颠倒的糖折叠分开了。S- 转向模体是在核糖体环 E 模体和八叠球菌 - 蓖麻毒素（sarcin-ricin）环中被找到的（见 8.3 节）。另一个重要的单链 RNA 模体是 U- 转向，这是在第一个和第二个核苷酸之间的骨架的突然转弯，紧接着是第二个和第三个核苷酸非常特殊的堆积。在第一个和第三个残基之间的氢键经常可以使模体稳定。U- 转向是 GNRA 环的典型特征，也在 tRNA 的 TΨC 环中被发现。

3.4.7.1 GNRA 四环结构（GNRA tetraloop）

四环结构（tetraloop）是在 RNA 结构中非常常见的模体，一个特别著名的发夹环的例子是 GNRA[a] 环模体，它在很多 RNA 茎环区域闭合发夹。在 SCOR 数据库列表中，有超过 600 个关于 GNRA 模体的例子。GNRA 环通常会采用特殊的三维结构，被称为 GNRA 折叠（GNRA fold），在 5′ 堆积端有一个碱基，而在 3′ 堆积端含有三个碱基。这一折叠含有一个 U- 转向，通过一个修剪过的非规则碱基对来稳定。GNRA 环模体经常由一个 CG Watson-Crick 碱基对来闭合。

GNRA 折叠在环上会出现 4、5 或者 6 个核苷酸。一个五环（pentaloop）的例子显示了 GNRA 折叠（共有序列 GNRNA）是 GAAAA。在这种情况下，第四个碱基是在环外侧的，而剩下的碱基采取与 GNRA 四环相同的空间位置（图 3.36）。序列 UMAC 的四环家族会形成相同的折叠（M=[AC]）。

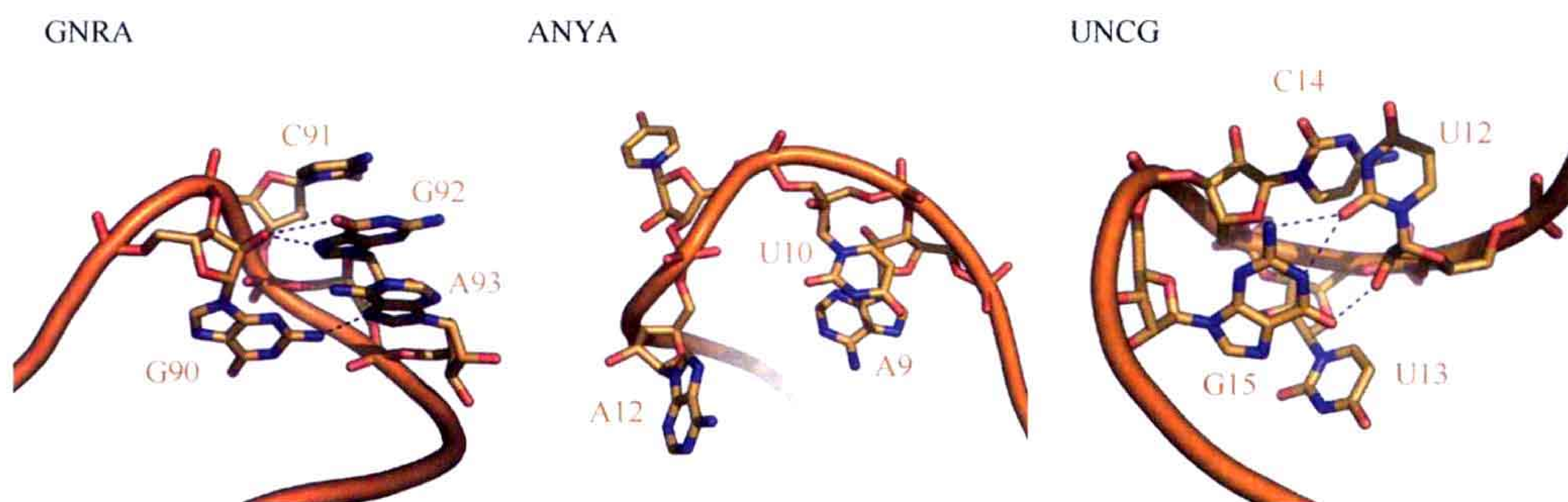


图 3.36 各种四环折叠的三维结构。左：5S rRNA 中的 GNRA 环（PDB：1JJ2）。第一个 G 在环中通过与第四个碱基形成的氢键稳定了环。中：MS2-RNA 复合物的 ANYA 环（PDB：1DZS）。碱基一和二形成了一个堆积相互作用，而环中的碱基三和四在环的外侧，可以很容易地和其他的片段相互作用。右：16S rRNA 的 UNCG 四环结构（PDB：1BYJ）。四环中的第一个 U 和最后一个 G 通过氢键相互作用，而环中的碱基一和二则形成堆积相互作用。环中的第三个碱基可以和其他的片段进行相互作用。

与 GNRA 类似，UNCG 模体也是一个在核糖体和一些其他的功能性 RNA 中可以找到的稳定的四环结构。一般的 UNCG 折叠（UNCG fold），在 5′ 堆积端含有 U 和 C 碱基，在 3′ 堆积端含有 G 碱基，以及在环外含有 N 碱基。同样的折叠也在 GUUA 四环中被观察到过。

a R 表示嘌呤；N 表示任何一个，Y 表示嘧啶。

另一个著名的四环模体是 ANYA2，在 MS2 病毒 RNA 外包蛋白结合的发夹环中非常常见（见第十五章）。

3.4.7.2　内环（internal loop）

通常，内环会与两个整齐的 A 型螺旋毗邻，环内的碱基可以不配对，它们经常参与规则的碱基配对（图 3.37）。紧靠的大量的水分子稳定了这一环区，而这些与水杂合的氢键加宽了深的小沟。

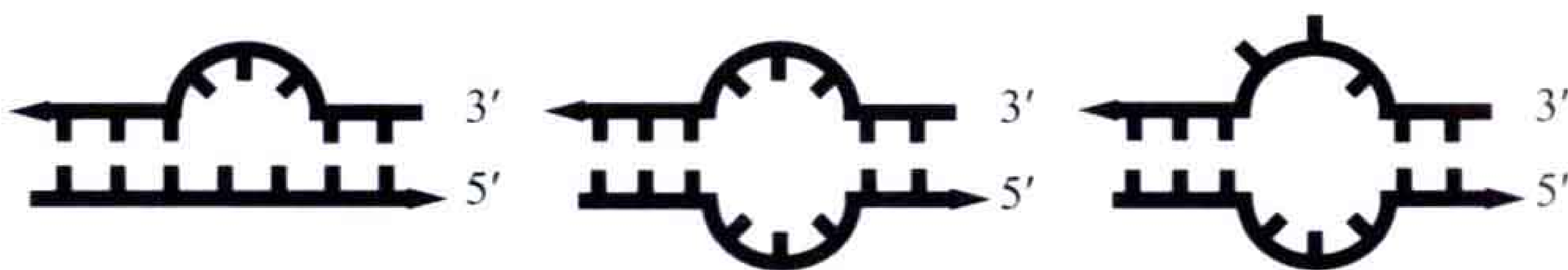


图 3.37　内部 RNA 环的各种例子。左：一个三碱基凸起，中：含有环内侧碱基的对称内部环；右：含有环外侧碱基的不对称内部环。

内环通常是蛋白质的结合位点，如 HIV-1 Rev 蛋白结合在 10 个核苷酸反对称的内环上。

内环可以根据毗邻螺旋的碱基配对的性质及环自身碱基的堆积来分类。显而易见，这些可以导致很多不同的环构象（表 3.2）。

表 3.2　根据 SCOR 分类的内环类别

内部环	亚类
堆积的、完全配对的非 Watson-Crick 双链	9
堆积的，一个碱基没有配对，左右两边都有碱基	2
环外碱基	2
含有三倍体碱基的环	3
含有二核苷酸平台的环	3
含有反方向糖苷键的环	4
含有不配对、不堆积碱基的环，碱基在环内侧	24
含有十字堆积的环	6
含有交错堆积碱基的环	n/a
含有不规则堆积的环	n/a
含有外部堆积碱基的环	2
扭结转向	n/a
螺旋转弯	4
含有两个独立堆积的环	n/a
S 转向	2
含有潜在配对可能碱基的环	n/a

注：这些亚基的出现不明，模体可以在很多细节上进行分类。

关于内环的一个得到广泛关注的例子是环 E 模体（loop E motif），其第一次被发现是在 20 世纪 80 年代，当时在真核 5S rRNA 和 PSTV 病毒 RNA 中发现相似及保守的内环结构，当它们处在紫外光下时，会共用一个奇怪的部分来形成交叉结合（图 3.38）。后来，环 E 模体在核糖体 23S rRNA 中的 sarcin-ricin 环中也被发现了，包含在延伸因子 EF-Tu 和 EF-G 结合位点中。在大肠杆菌 5S rRNA 中，环 E 是已知的用来结合核糖

体蛋白 L25 的特异性结合位点（图 3.39）。因此表面上这个模体是一个重要的活性位点和分子识别区域，通过高分辨率的 NMR 和晶体学结构，5S rRNA 和 sarcin-ricin 环都已经被研究得很透彻了。

图 3.38 共有的环 E 家族模体。左边的二级结构图中的圆圈指出了非 Watson-Crick 碱基配对。

图 3.39 上：*H. marismortui* 的 5S rRNA 的二级结构，在一个大核糖体亚基的晶体结构中观察到的；下：5S rRNA 的三维结构的图解表述，与上图的方向是一致的。

环 E 是一个非对称的内环，它具有高保守的 7 个非 Watson-Crick 碱基配对堆积的特点。它的氢键模式也特别保守。第一个碱基对是一个 AG 共用对，还有一个 UA 反 Hoogsteen 对，接着有一个 G 碱基凸起，最后由一个反 Hoogsteen AA 碱基对结束，形成了一个局部平行的骨架构型（图 3.38）。AU 颠倒 Hoogsteen 碱基对也在环 E 中被观察到，它是在核糖体 RNA 中观察到的除正常 Watson-Crick 碱基配对以外最多 AU 相互作用的例子。

3.4.7.3　凸起 (bulge)

在另一侧是规则螺旋的区域插入一定数量的碱基后形成的结构模体，被称为凸起，但是事实上这只是内环结构的一个特殊情况（图 3.40）。一个凸起插入 RNA 双螺旋的一个转弯中，这是一个重要的结构组件，因为它有确定茎环结构取向的能力，它也是通过这样的贡献来影响整个分子的三级结构折叠的。



图 3.40　一个碱基在环内的凸起模体。

在 RNA 凸起中未成对的核苷酸可以在环的外侧，也可以在堆积区的内侧。如果在环的外侧，核苷酸经常是形成三级结构相互作用的位点，或者是蛋白质的特异性识别位点。图 3.40 显示了一个单碱基凸起，但是在凸起中可以有不止一个的碱基。

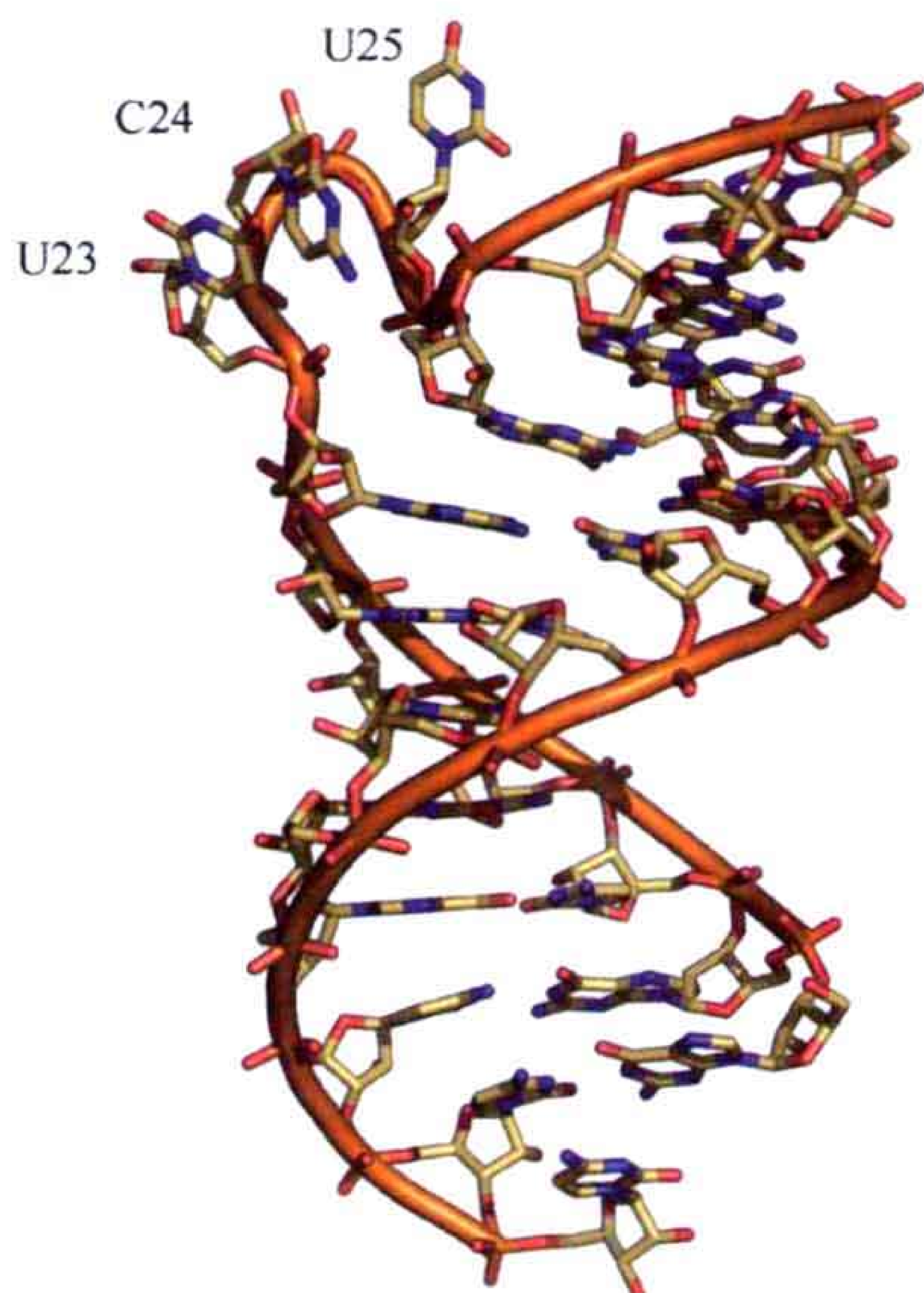


图 3.41　TAR 茎环结构中的凸起特写（PDB：397D）。

一个具体的、已经被研究得很透彻的 RNA 凸起是存在于 HIV-1 的超激活（trans-activation，TAR）响应区域（图 3.41）。这是一个含有 59 个核苷酸的 RNA 茎环结构，存在于所有的 HIV-1 mRNA 中的 5′ 非编码区域，它在调控 HIV 基因表达中扮演了重要的角色。这个 TAR 区域与病毒 Tat 蛋白相互作用，来激活病毒的基因表达，而且这对病毒复制来说是必需的。这里有实验证据显示 Tat 结合是直接指向 TAR 片段的，而且这一结合位点是与核苷酸序列相互独立的，它取决于茎上部的完整性，包含一个 3- 核苷酸的凸起。

3.4.7.4　连接（junction）

连接即 RNA 连接两个或更多茎结构的区域。在各个茎结构之间的单链可以有大于等于 0 的长度，这些单链被称为连接（linking）或者联合（joining）区域（图 3.42）。三茎（three-stem）结构连接是在稳定的 RNA 中最常见的，如核糖体 RNA、病毒 RNA 及核糖酶。在后面的章节中我们将进一步讨论锤头状的核糖酶，它是一个真正的三茎连接。四茎（four-stem）连接——有时候被称为十字形连接（cruciform junction），也

是一种比较常见的结构，如在 tRNA 中。

三茎连接的拓扑结构已经被描述得非常清楚了。两个螺旋同轴堆积的三茎连接可以分成三个家族，根据它们的链接区域的长度，这再一次显示了三条茎的相关方向。如果链接比较长，螺旋则可以相对于彼此旋转，使得两茎相互堆积，因此链接区域的长度在一些情况下是非常重要的，例如，如果链接比较长，因为螺旋允许自身相对于其他的螺旋转动，则茎两两之间缠绕。在图 3.43 中，展示了三个家族的三茎连接结构的图解描述。在家族 A 中，链接区域 J31 比 J23 要短，螺旋 P3 近似垂直于 P1-P2 堆积轴。在家族 B 中，链接 J31 和 J23 一样长，螺旋 P2 和 P3 相互缠绕。家族 A 和家族 B 连接可以在核糖体 23S 与 16S rRNA 中被找到。最后是家族 C，链接 J31 比 J23 要长，螺旋 P3 与螺旋 P1 平行缠绕，以及链接 J31，大部分是发夹结构，与 P2 的浅沟具有大量的相互作用。家族 C 连接可以在各种各样的结构 RNA 中被找到，如垂头状核糖体、自黏结的 P4-P6 内含子（intron）区域，以及信号识别特定的区域 S 和 G 核糖转换（riboswitch）

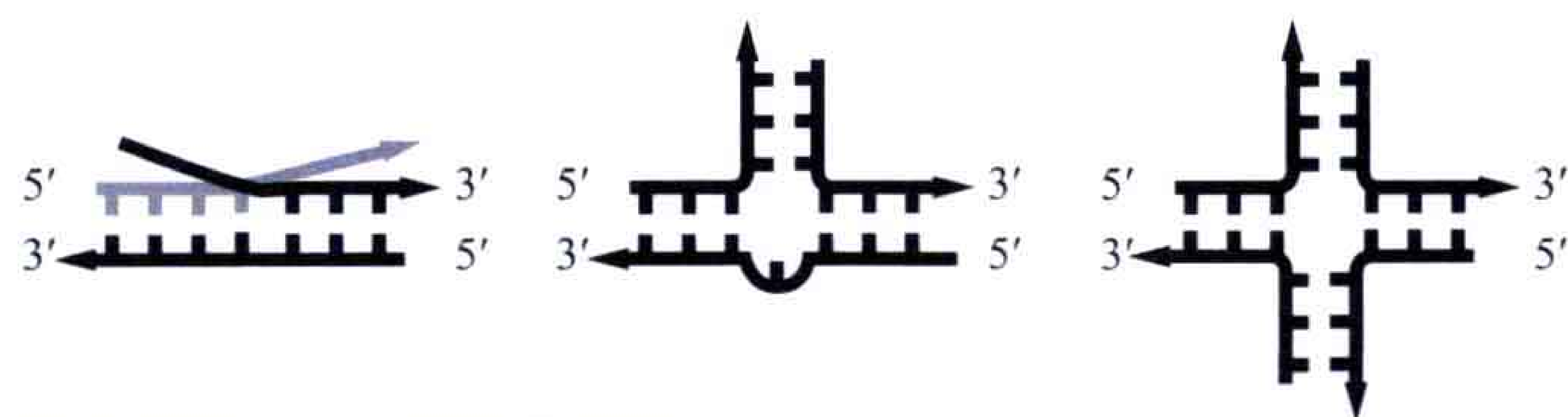


图 3.42　二茎、三茎和四茎 RNA 连接的图解。

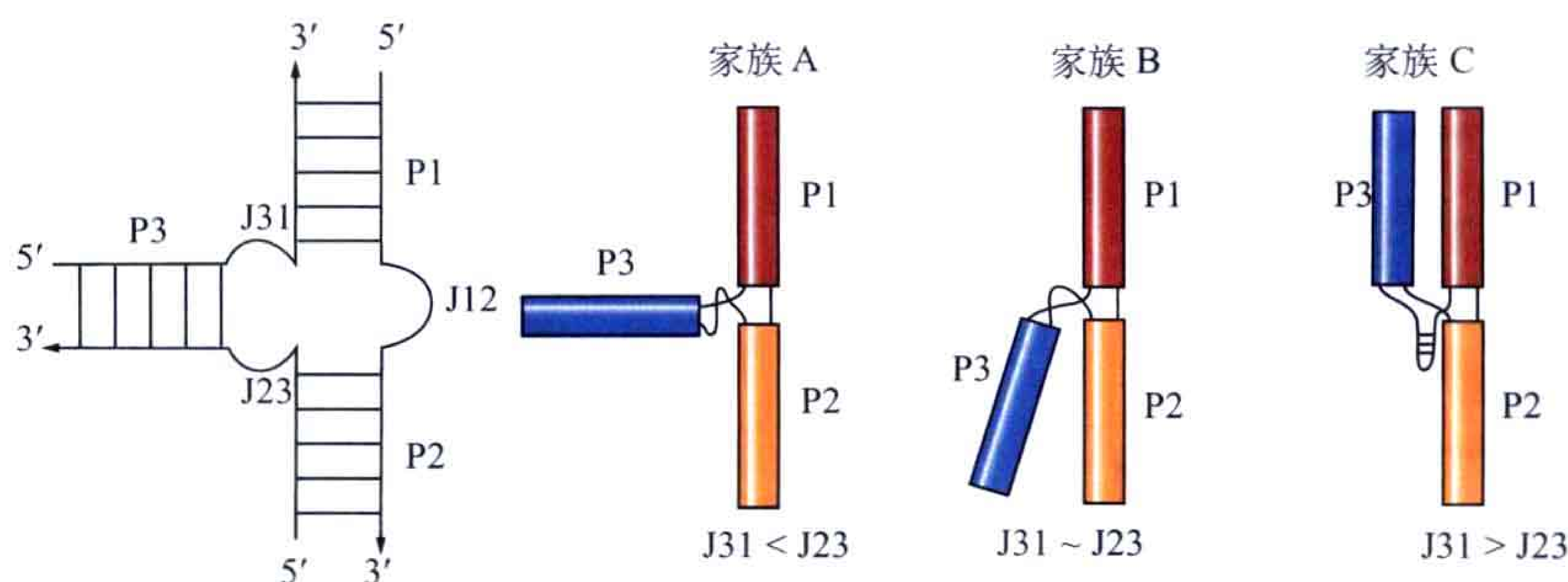


图 3.43　左：三茎连接的各部分命名；右：三种方式连接的三个家族的图示。

3.4.7.5　K- 转向 (K-turn)

对 *Haloarcula marismortui* 的 50S 核糖体颗粒的三维结构的分析，使得一个小 RNA 模体被发现，这一模体被称为扭结 - 转向（kink-turn），或者 K- 转向。K 转向是一个两条链的、含有大约 15 个残基的环 - 螺旋模体结构（图 3.44）。这是一个反对称的内部环，在糖 - 磷酸骨架上具有一个扭结特征，这是因为在 RNA 螺旋上有一个急转弯。弯曲出现在浅沟边，使两个毗邻的浅沟对接到了一起。其中一个螺旋，被称为 C- 茎（规则茎），只含有 Watson-Crick 碱基对；而另一个螺旋含有不规则的碱基配对，因此被称为 NC- 茎。

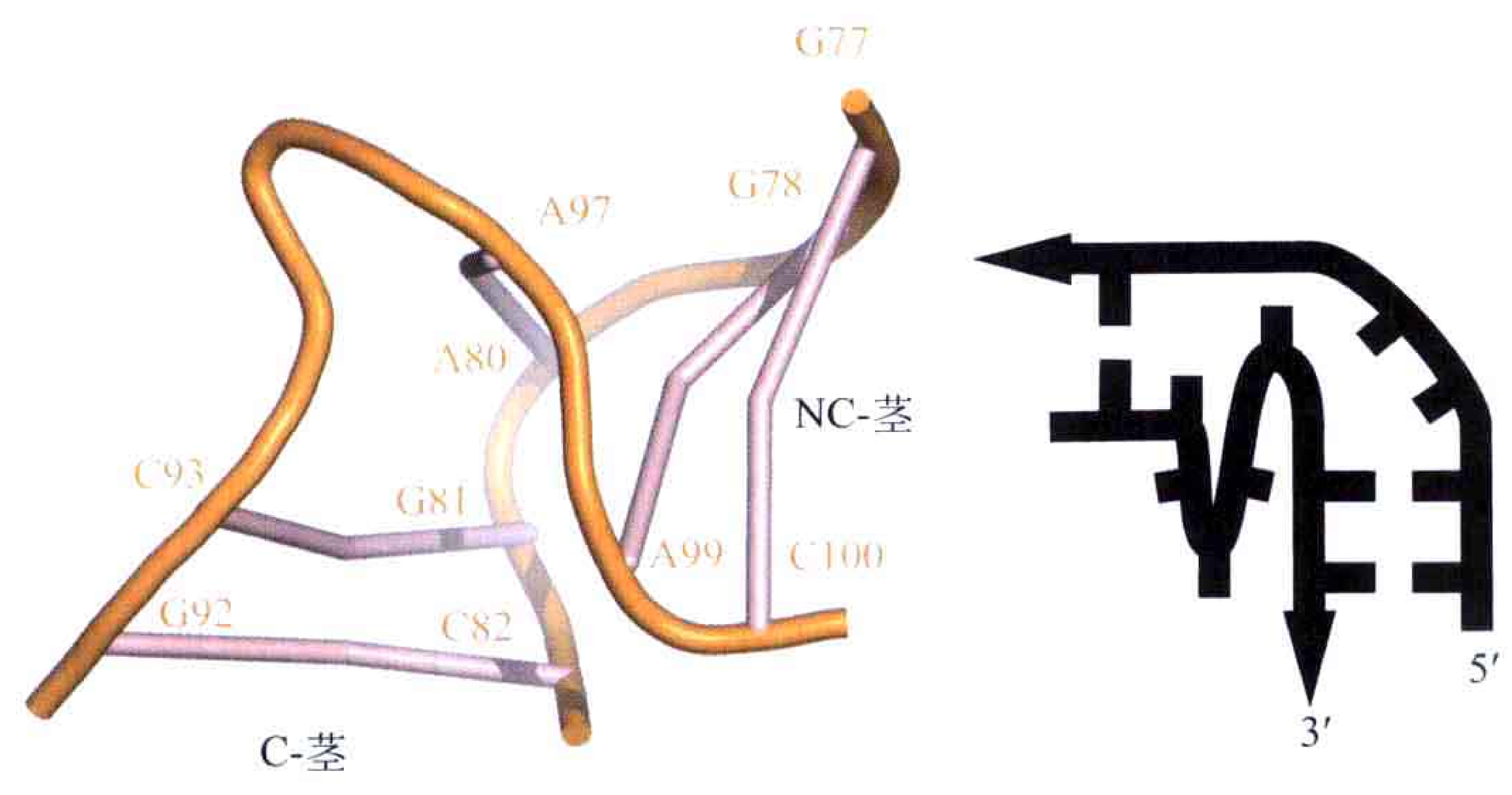


图 3.44　左：*H. marismortui* 23S rRNA 的螺旋 7 中的 K- 转向图示；右：K- 转向的二级结构图谱。

这两个茎通过一系列的堆积相互作用被叠在一起，包括一个 A-minor（A- 次级）相互作用。未成对的环上核苷酸中一个显著的凸起，使这个模体成为核糖体蛋白识别的理想位点。在 *H. marismortui* 50S rRNA 中有 6 个 K- 转向，而在 *T. thermophilus* 30S rRNA 中有两个，U4 snRNA 和 L30e mRNA 前体中各有一个。在很多 RNA 中 K- 转向序列具有一个共有序列模体，包括 L10 mRNA 的 5′ -UTR，大肠杆菌 23S RNA 中螺旋 78 和人类核糖核酸酶 MRP。K- 转向一致地在上面部分含有两个 CG 碱基对，在下面部分含有两个 AG 碱基对，一般随后是两个 Watson-Crick 碱基对。扭结一般在一条链上含有三个碱基，它们不是成对的碱基。

K- 转向未成对环上的核苷酸中有一个显著的突出，这使得 K- 转向成为重要的 RNA 识别蛋白模体：在 23S rRNA 的 6 个 K- 转向中至少有 5 个与一个核糖体蛋白具有显著的相互作用。

3.4.7.6　括号记号（bracket notation）

RNA 二级结构可以用被称为括号记号的方法来表示，即核酸序列可以用等长的含有点和配对括号的一串符号来表示。未成对的碱基用点号（有时候用冒号）来表示，而碱基对在碱基 *i* 和 *j* 之间的 *i* 位置用一个“（”、*j* 位置用一个“）”来表示。这样二级结构就成了

(((...((((.....))))..)))

它表示如图 3.45 所示的茎环结构。

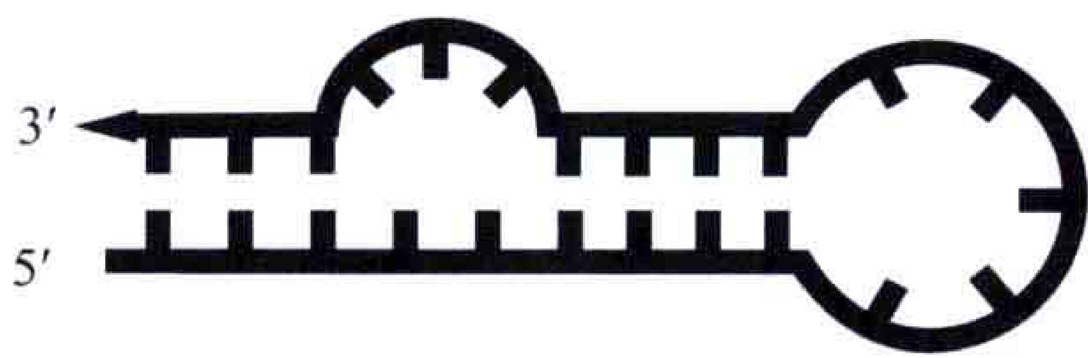


图 3.45　与上文中括号记号一致的茎环结构。

当环处于一个序列的中间位置时，括号记号方法是最方便阅读的方法。在上面的例子中，茎 - 五环结构的四个碱基对在茎中通过字符串 ((((.....)))) 给出，同时可以向任

何一边延伸结构。内部的环，在一条链上含有三个碱基，在另一条链上则含有两个碱基，从记号中可以看到，它们分得非常开。对于较短的序列来说，括号记号是非常容易被理解的；但是对于大结构而言，记号则变得相对复杂，同时它也非常适合作为计算机程序的输入文件。

3.4.7.7 RNA 假结（pseudoknot）

假结是这样的一个结构，在 RNA 二级结构中包含两个茎环结构，而且第一个茎环结构的环是第二个的茎。假结于 1982 年第一次在芜菁黄花叶病毒中被发现的，但是直到在很多 RNA 结构中被发现，大部分假结依然是来自病毒的。假结会折叠成更加紧凑的三维构型但却不是真正的拓扑结。事实上，这更多是在三级结构层面上相互作用形成了双螺旋。

假结包括两个相互作用的茎环区域，根据大部分二级结构的预测算法，不依赖于茎环区域的分层嵌套假设。然而，只需做一点细微的改动，括号记号法就能用来描述假结了。在图 3.46 中，如果圆括号被用来描述其中一个茎环结构，而方括号被用来描述另一个，这样记号结果将变成：

(((.((((........[[[[[[[[[))))).))) - - - .....]].]]]]]]

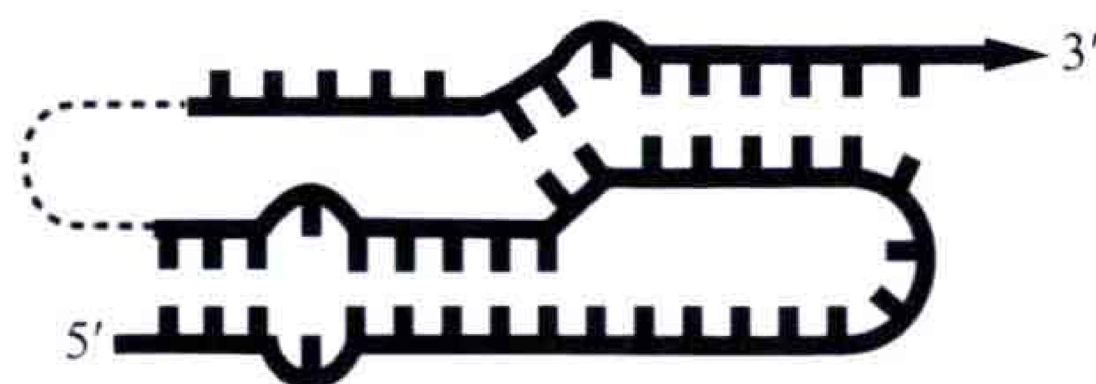


图 3.46　一个假结的图示结构。假结包含有两个相互作用的茎环结构，因此一个茎环的环是另一个的茎。

图 3.47 展示了一个含有适配子 [b] 的假结结构，它是由一个随机序列分子的组合分离得到的。这一适配子相对于它的结合生物素具有选择性。生物素辅助因子结合在一个口袋中，在假结中堆积的螺旋之间的界面上。生物素结合蛋白，如抗生物素蛋白，把生物素深埋在疏水核心中，但是适配子依赖于水合镁离子和水分子来结合它的配体。这一策略展示了不同途径蛋白质和 RNA 分子识别的基础。

3.4.7.8 A- 小模体（A-minor motif）

距离较远的核苷酸碱基对之间的相互作用对 RNA 三级结构的稳定具有显著的贡献。众所周知，腺苷酸是一个非常常见的核苷酸，主要在外侧规则螺旋的保守位置出现。对于 50S 核糖体亚基中 rRNA 的检测显示，大部分的 A 碱基被卷入三级结构的相互作用中，且这些作用是根据一个有限的模式形成的，此模式被称为 A- 小模体（图 3.48）。

---

b 适配子是这样一种核酸，由物种进化中通过体外选择的重复循环结合到不同的分子目标上产生的，如小分子、蛋白质或者核酸。

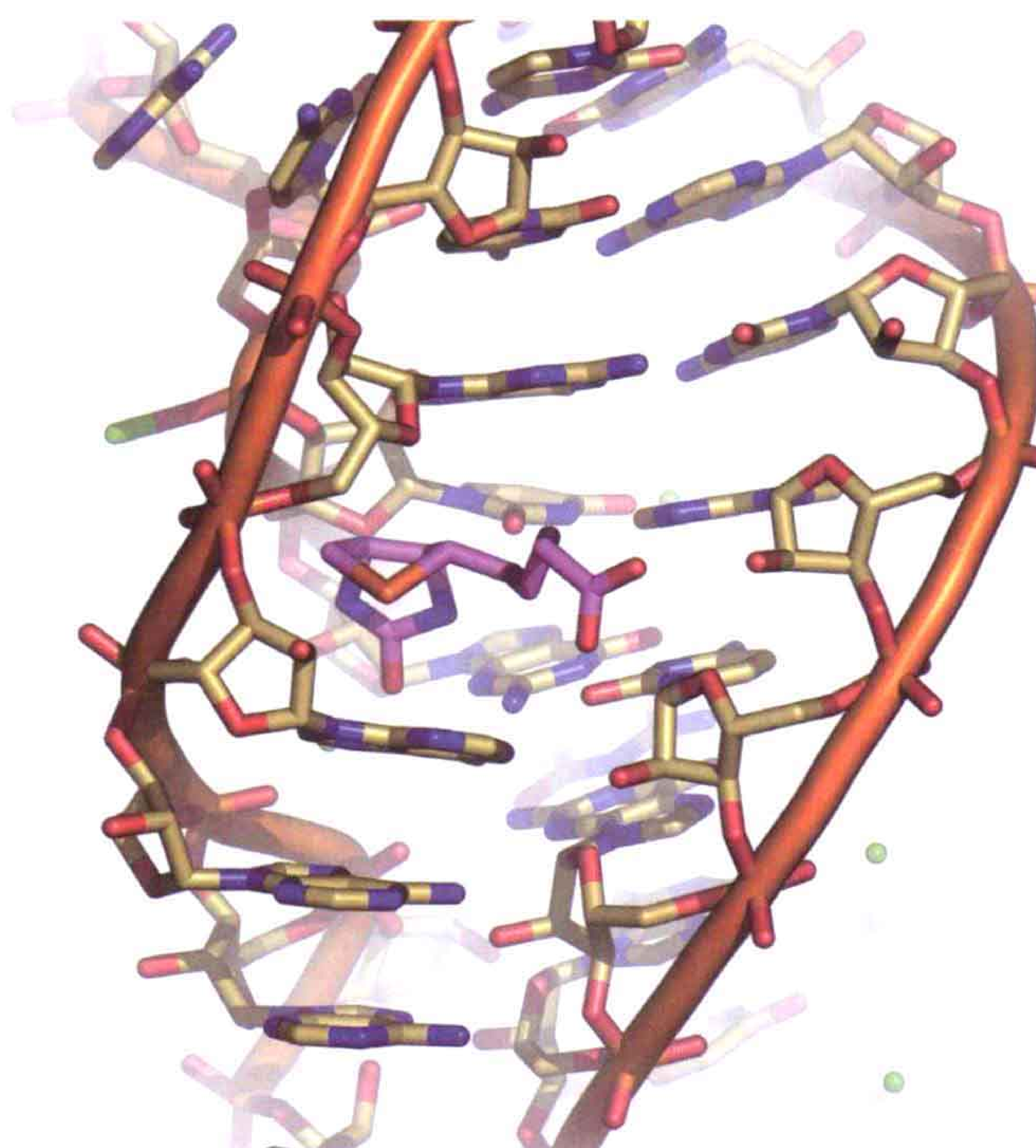


图 3.47　在假结中生物素结合位点的细节。生物素被显示为紫色（PDB：1F27）。

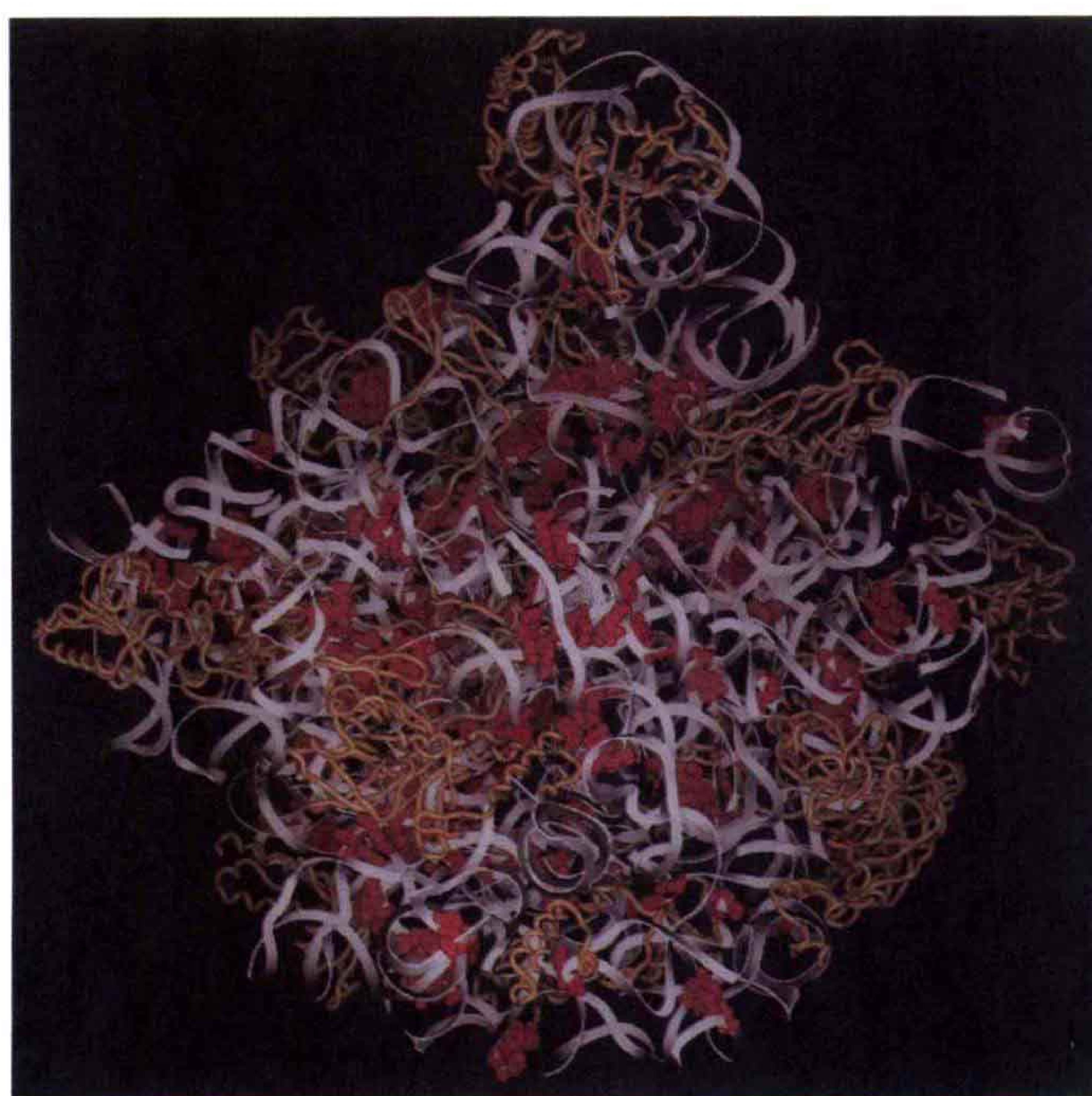


图 3.48　*H. marismortui* 中 50S 核糖体亚基的结构，显示了形成 A- 次级相互作用的 186 个腺苷酸。

确实，在 *H. marismortui* 23S rRNA 中 26% 的腺苷酸碱基（其中 64% 的碱基的保守性大于 95%）与 RNA 浅（小）沟通过它们的 N1-C2-N3 边缘相互作用，它们很光滑是因为缺少了其他碱基环外的原子。

迄今的研究表明，在 RNA 结构双螺旋的堆积中，A- 小模体是主要的三级连接，因此 A- 小模体可能是形成 RNA 三级结构最重要的结构组件。

一共有 4 个不同的 A- 小模体，其 A 碱基上 O2′ 和 N3 原子位置的不同，这与受体螺旋中碱基对的 O2′ 原子有关（图 3.49）。在类型 I 模体中，A 碱基的 O2′ 和 N3 原子都在受体碱基对浅沟的内侧（图 3.49）。这一排列使得腺苷酸更加完美地匹配沟槽，也更大限度地使得氢键可以形成。

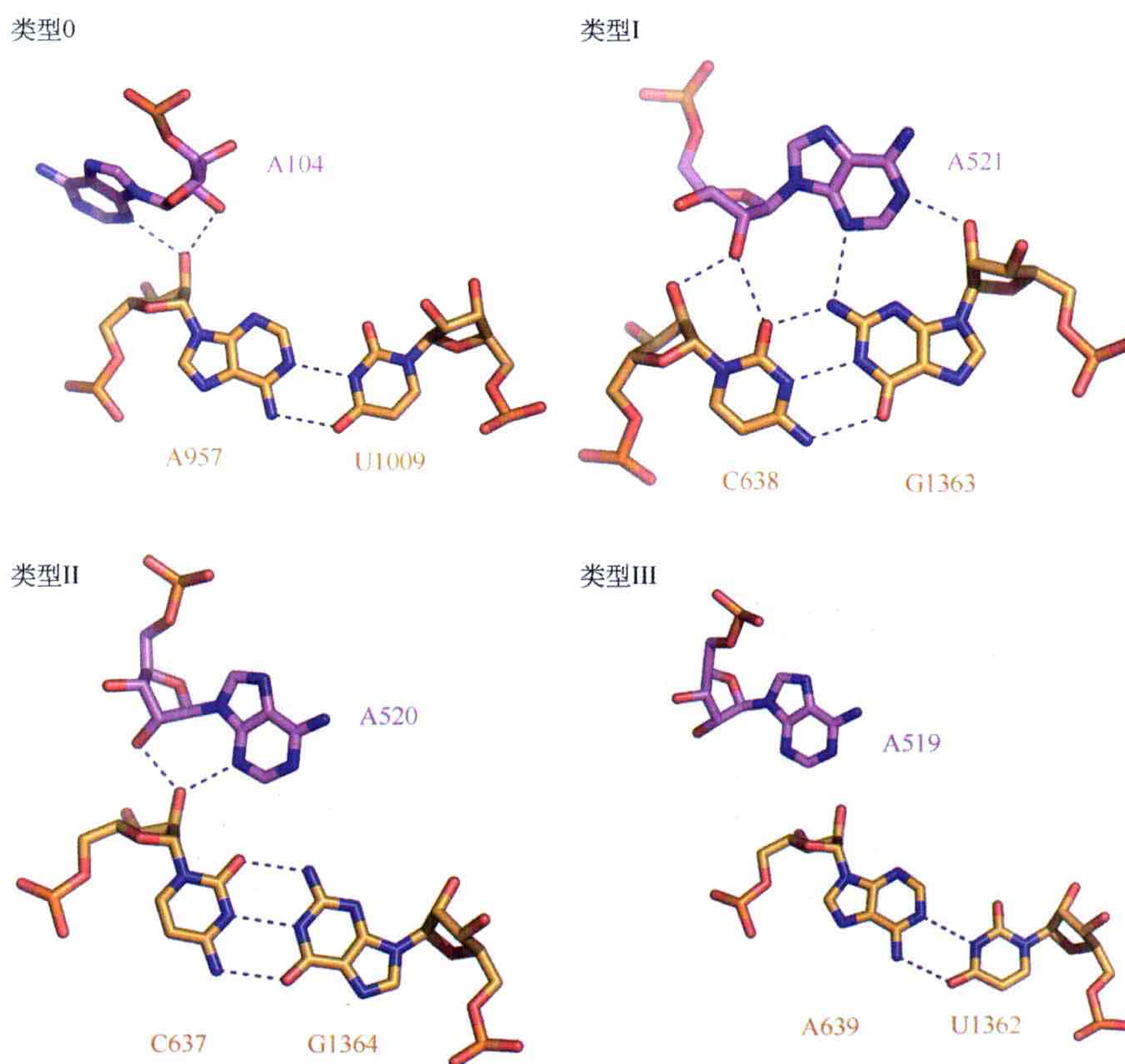


图 3.49　4 个 A- 小模体相互作用类型的例子，来自 *H. marismortui* 23S rRNA。

在类型 II 模体中，A 碱基的 O2′ 原子在浅沟的外侧，而 N3 在内侧，同时与螺旋糖磷酸骨架上的 2′ 羟基基团相互作用（图 3.49）。类型 III 模体的特点是，A 碱基的位置使得 A 碱基上的 O2′ 和 N3 原子都处在浅沟的外侧（图 3.49）。在第四种类型中，很少被提到的类型 0 模体，其 A 碱基的核糖与受体螺旋骨架上的核糖相互作用（图 3.49）。类型 0 相互作用不是碱基特异性的，因为它是插入残基的核糖，填满了浅沟，而事实上，类型 III 也不是碱基特异性的。然而，在类型 0 和类型 III 相互作用中，当碱基是腺嘌呤时，

插入碱基和受体螺旋之间的接触是最优化的。A- 小模体中插入的 A 碱基具有对 GC 受体碱基对强烈的偏好，它在外形和氢键模式上是最佳的互补序列。

3.4.8 RNA 是遗传信息的载体

RNA 具有一个重要的角色，在染色体到核糖体的遗传信息传递中，可以作为传真，或是信使。信使 RNA（mRNA）是一个这样的 RNA 分子，它编码了一个单基因或者转录单元的拷贝。mRNA 是第一个通过 RNA 聚合酶从一条 DNA 模板上被转录下来的（第七章），它是为核糖体运载编码信息的，核糖体是蛋白质合成的场所（第八章）。在一些病毒中，单或双链的 DNA 组成了病毒基因组（第十五章）。

在 5′ 端（“头”端），真核生物信使 RNA 包含一个甲基化鸟苷 (m^7G) 帽、一个修饰了的鸟嘌呤核苷酸，其在转录的同时被加到了 mRNA 上（图 3.50）。这一帽子的出现对于 mRNA 被核糖体识别是至关重要的，另外，它还保护 mRNA 不被核糖核酸酶降解。在 3′ 端，聚腺苷酸聚合酶会加上一段腺嘌呤核苷酸序列形成带尾巴的 mRNA 前体。这一 poly（A）尾有时可以达到数百个核苷酸那么长。

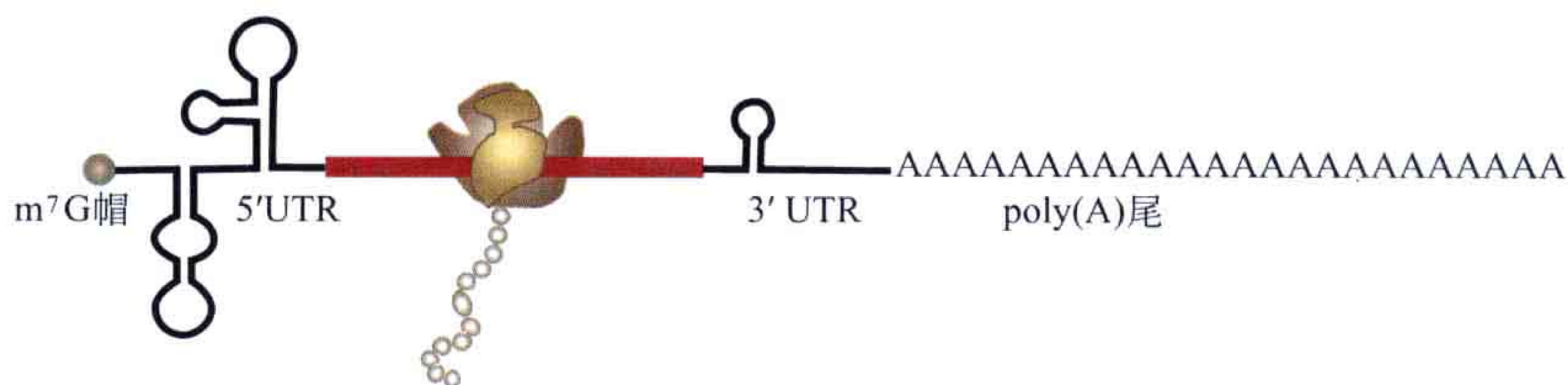


图 3.50 mRNA 的表述，它展示了 5′ 帽子、编码区域（红色），以及 5′ 和 3′ 端 UTR。

除了编码区域以外，mRNA 还包含一些不会被翻译成蛋白质的区域，其在 mRNA 稳定、mRNA 的定位和转移效率方面具有很重要的作用。这些不被翻译的区域（UTR）会出现在开始编码的前面（5′ UTR）和终止密码子的后面（3′ UTR）。它们含有界限清晰的二级结构区域。结合到 UTR 区域的蛋白质可以影响核糖体结合 mRNA 的能力，这样可以影响转运的效率。在 5′ UTR 区域的 mRNA 二级结构被融合了起始因子 eIF4A 和 eIF4B，这个过程需要 ATP 水解作用。3′ UTR 被认为是与 mRNA 在细胞内定位有关的，因此编码的蛋白质可以被转运到需要这个蛋白质的细胞的各个部分。mRNA 的编码区域可能同样含有二级结构元素，而事实上被确认的是，在 mRNA 中具有二级结构引导转运控制。

因此，mRNA 组成了一个可被理解的情报模块，不只是含有供合成编码蛋白质的谱图，还含有通信规则的信息，它是决定 mRNA 分子自身命运的信息。

3.4.9 RNA 具有重要的结构角色

RNA 的结构角色通过核糖体 RNA（rRNA）来体现，它形成了核糖体的大部分结构，

核糖体是合成蛋白质的场所（见 8.3 节）。核糖体（rRNA）是在生命的演化过程中最保守的基因，因此，rRNA 序列经常被用来作为分类学上的分类依据，来估计物种的分化时间。

70S 细菌核糖体的 X 射线结构展示了一个复杂的核糖体 RNA 结构。最初，rRNA 被认为仅仅是使核糖体蛋白附着的骨架而已，但是这些图片很快就被发现是不对的。现在知道，核糖体 RNA 与催化关系非常密切。一些核糖体蛋白以穿透 rRNA 结构的方式防止自由的折叠。尽管很小的 rRNA 在溶液中可以形成更加一致的三维结构，但是观察到的实验数据证明核糖体蛋白参与了 rRNA 的正确折叠，至少一部分蛋白质参与了 rRNA 的正确折叠。

一个 17 个碱基序列的环（原核生物中是 15 个碱基），被称为八叠球菌 - 蓖麻毒素（sarcin-ricin）环，是最高保守性的核糖体 RNA 之一。真菌的毒性酶八叠球菌在环上分解了一个磷酸键，蓖麻毒素（一个来自蓖麻子的毒素）在模体上脱掉了一个重要的腺嘌呤。八叠球菌和蓖麻毒素这样的作用破坏了 sarcin-ricin 环，防止了其与延伸因子的结合，因此这是核糖体的固有功能。

RNA 的结构功能还可以被归因于 UTR，这已经在上面讨论过了。UTR 在不同 mRNA 的多序列比对显示了其非常好的二级序列保守性，这再一次证明了其自身的三维结构保守性。因此，mRNA 的三维结构隐藏了对于蛋白质表达调节与控制的玄机。

3.4.10 RNA 识别

小 RNA（miRNA）是最近才被发现的一类 RNA，是 21~22 个核苷酸长的来源于前体细胞的 RNA；miRNA 是在基因组中编码前体细胞的茎环结构 RNA。成熟 miRNA 的功能在大复合物中是对特定目标 mRNA 起负调控作用。它是目标序列的互补碱基对，经常出现在 3′ -UTR 区域，在翻译被抑制时出现，这是因为核糖体无法结合 mRNA。这一类型 mRNA 调控的发起者被称为核糖调控子（riboregulator）。

与 miRNA 活性相关的是 RNA 沉默的机制，其通过小干扰 RNA（siRNA）来调节。siRNA 是一个含有 21~25 个核苷酸的双链 RNA 分子，由双链 RNA 通过核糖核酸酶Ⅲ酶切产生。就像 miRNA，siRNA 也具有转录后使基因沉默（PTGS）的能力，有时还被称为小瞬间 RNA（small temporal RNA，stRNA），表明它们只是短暂地起到调控作用。

RNA 干扰，或者 RNA 沉默，是通过 mRNA 降解或者翻译抑制来调控基因表达的。简而言之，siRNA 和 miRNA 之间的不同可以被归因于短 RNA 和它的目标之间互补的程度。miRNA 是由 RNA 的酶切过程产生的，所以和二聚体的相反链不匹配（图 3.51 左），而 siRNA 是由规则的双链 RNA 产生的（图 3.51 右）。

大的核糖核蛋白复合物介导了这些 RNA 的沉默过程。在 RNA 的干扰过程中，这一复合物被称为 RISC（就是 RNA 诱导沉默复合物）。Argonaute 蛋白就是 RISC 复合物的一个具有催化能力的组成部分。Argonaute 蛋白会结合 siRNA 片段，然后对与结合的 siRNA 片段具有互补序列的 mRNA 链产生核酸内切酶活性（图 3.51）。

RNA 干扰是一个细胞在很短暂的时间尺度上可以将一个特定的基因表达终止的机制。例如，一个来自 *C. elegans* 的经常被研究的 RNAi 体系，其 miRNA 可以控制胚胎后发育的时间和顺序，同时还可以调节蛋白质在一个精确的发育阶段表达。RNAi 的运用是一个有效的和经济的途径，其使得细胞实现调控，因为蛋白质机器可以被不同的 RNA 重复利用。RNA 干扰机制是一个非编码 RNA 成为一个分子开关组成部分的例子，可以识别特定基因并且对基因表达进行微调。

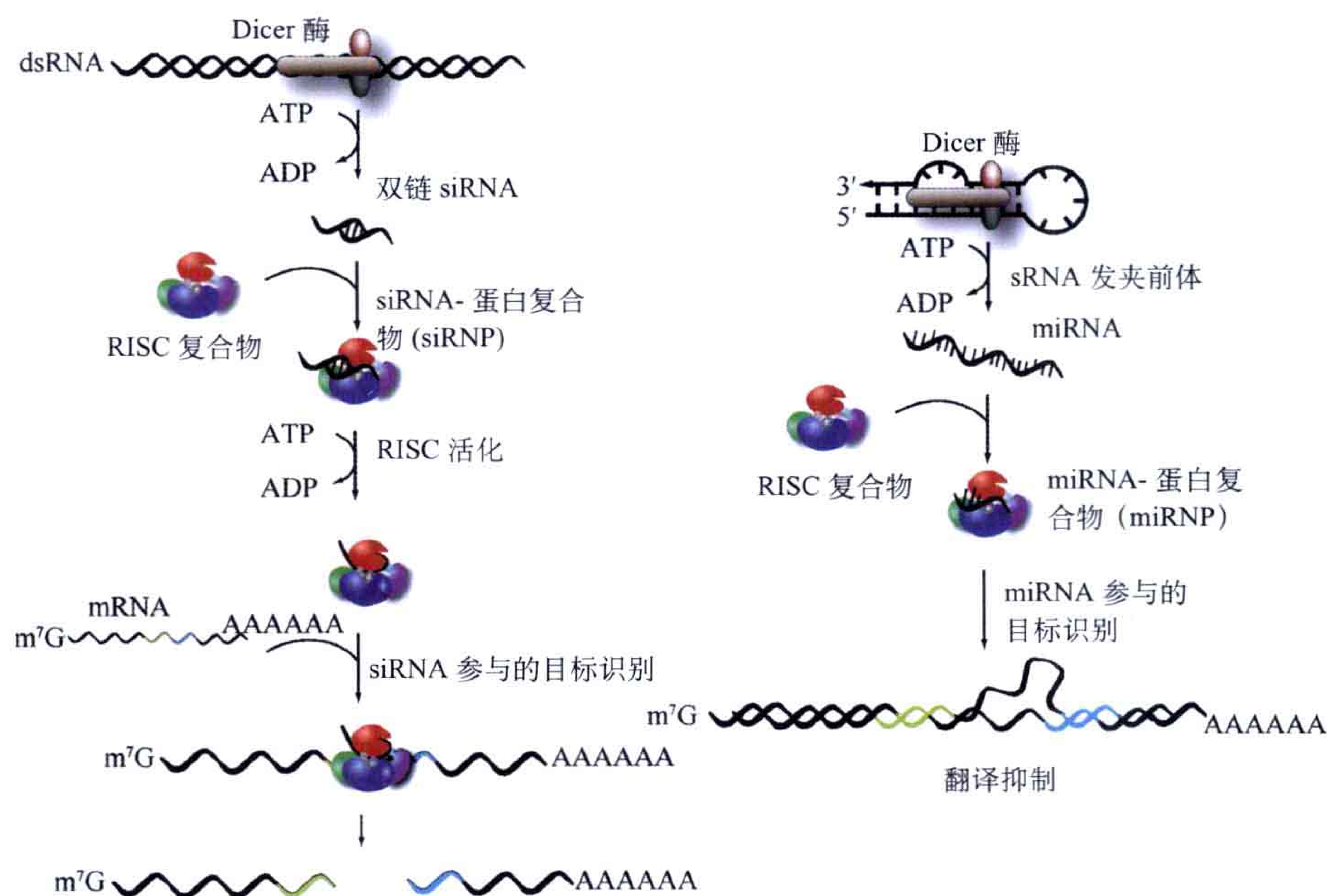


图 3.51　左：siRNA 参与的 mRNA 剪切。长的 dsRNA 可以通过与核糖核酸酶Ⅲ类似的酶切来加工形成 siRNA。加工过的双链 siRNA 解链分开，成为 RNA 诱导沉默复合物（RISC）的一部分。RISC 的蛋白质组分结合 siRNA 的一条单链，利用它来寻找可被识别的含有互补序列的目标 RNA。目标 RNA 识别位点的中央位置将被 RISC 中的核酸内切酶 Argonaute 砍断。右：小 RNA（miRNA）途径。长的、不匹配的发夹结构可以通过酶切加工产生单链 miRNA，它可以组成 miRNA- 蛋白复合物（miRNP）。miRNP 复合物可以利用 miRNA 的部分互补序列与目标 mRNA 配对，导致翻译抑制。

3.4.10.1　tRNA

另一类 RNA 是转运 RNA（tRNA），它们的最主要任务是识别。tRNA 属于小 RNA 类（73~93 个核苷酸），主要在翻译阶段起作用，它的主要任务是转运特定的氨基酸到核糖体蛋白合成位点增长的多肽链上。每种 tRNA 含有由三个碱基组成的反密码子，与特定 mRNA 上的编码正好匹配。在分子的末端，tRNA 承载了氨基酸，与它的 3′ 端共价连接，与特定的编码匹配。每一种 tRNA 只可以运载一种类别的氨基酸，但是因为基因编码是简并的，所以出现了多种 tRNA 可以运载一个氨基酸。唯一没有匹配 tRNA 的编码是终止密码子 UAG、UAA 和 UGA。tRNA 在翻译中的功能将在 8.3 节中更加详细地说明。

tRNA 的结构是 1974 年被两个独立的实验室解出的，一个是英国剑桥的 MRC，另一个是美国的 MIT。事实上，在 Crick 和 Watson 提出他们的 DNA 结构模型后，这是第一个核酸二聚体的晶体结构。

不同类别的 tRNA 具有非常相似的序列，在 3′ 端运载一个氨基酸的单链凸起往往是 CCA。tRNA 包含很多修饰的碱基，尤其当腺苷酸出现在反密码子的第一个位置时，它经常被修饰为肌苷（inosine，Ⅰ），它是缺少了氨基基团的嘌呤环。肌苷对其碱基配对伙伴的要求是不严格的，可以与 A、U 或者 C 配对，这也是为什么会有基因编码简并的原因。

tRNA 的二级结构是一个四茎连接结构，但是因为二级结构特征的出现，经常会被描述为“三叶草形”（“cloverleaf”）（图 3.52）。其中两个环根据它们包含的修饰碱基被命名为 D- 环（dihydro-U）和 TΨC 环（或者 T- 环）。

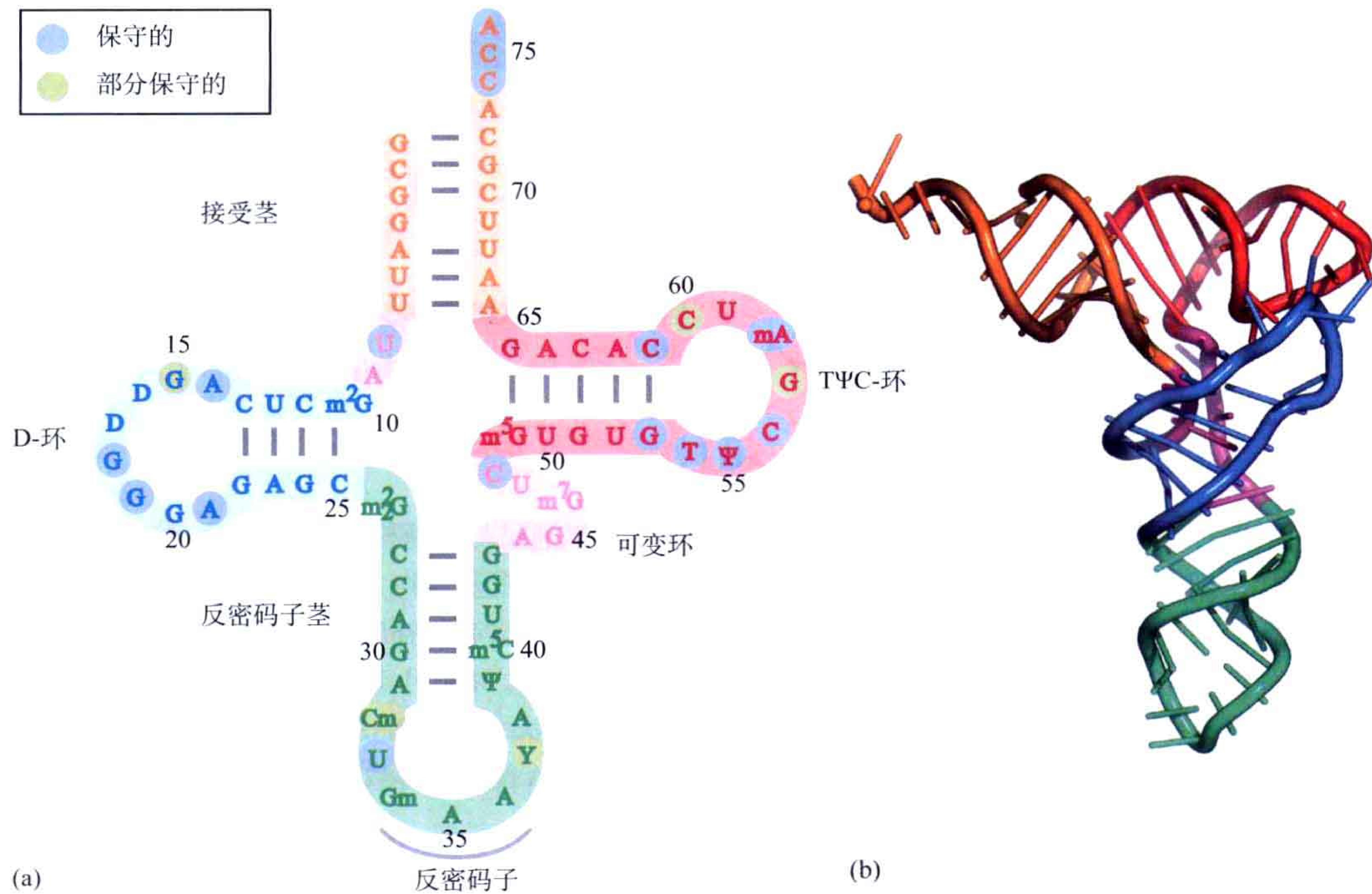


图 3.52　(a) 酵母中 $tRNA^{Phe}$ 的二级结构；(b) tRNA 分子的三维折叠图，与 (a) 中的颜色一一对应。

tRNA 的连接区中心，即非螺旋区域的三维结构是非常复杂的。骨架、碱基及核糖环的 2′ - 羟基基团，甚至是水分子和镁离子都参与了保持 tRNA 结构紧凑的、精密的相互作用网络。图 3.53 从概念上阐述了三叶草形二级结构是如何折叠成一个三维结构的。受体臂（acceptor stem）堆积到了 TΨC 茎，形成了一个同轴螺旋；同样，反密码子臂堆积到连接序列的上面，在可变的环和 D- 臂之间，形成了另一个近似完美的、同轴的螺旋。

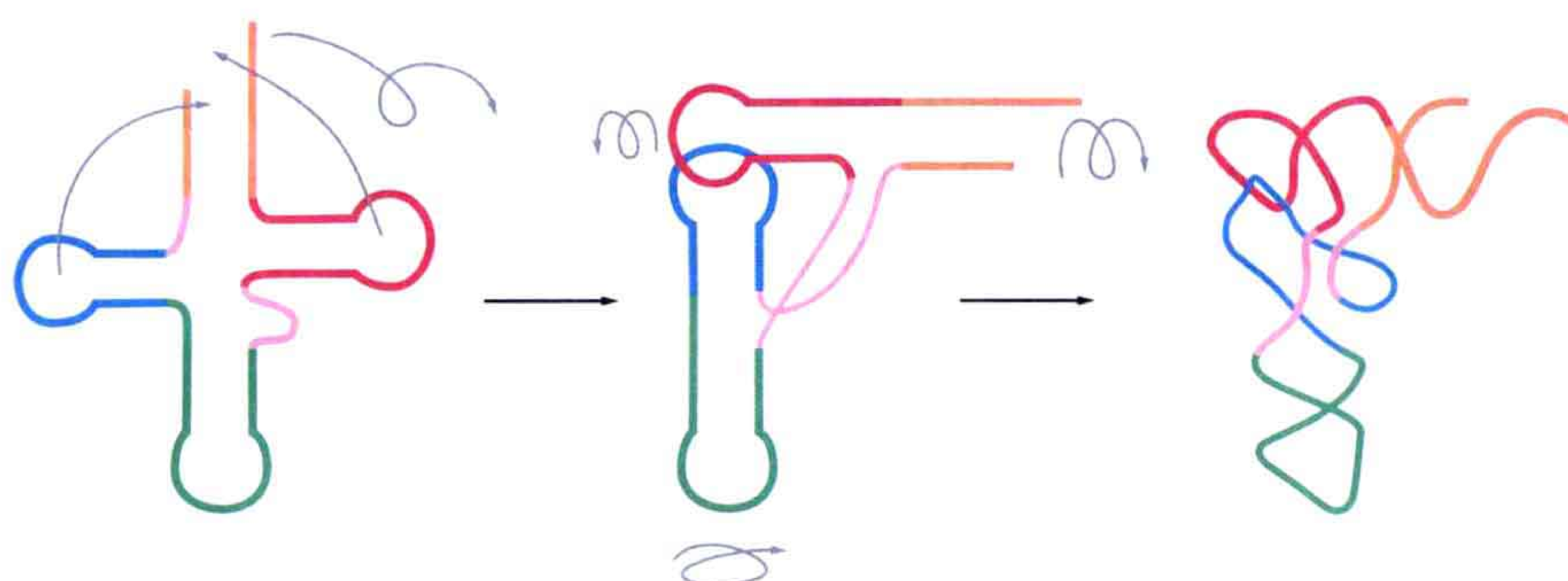


图 3.53　tRNA 链折叠的图解描述，很多保守残基是 tRNA 三级结构相互作用的保证。这里的颜色设置与图 3.52 相同。

在 TΨC 和 D- 环之间的三级结构层次的相互作用使得叶草形结构成为一个 L 形三级结构。T- 环上的核苷酸 G-57 和 A-58 与 D- 环上的 G-18 和 G-19 形成了堆积，C-56 与 G-19、G-15 与畸变环上的 C-48 分别形成了碱基对。这些相互作用有效地使分子的两部分扭结到了一起（图 3.54）。

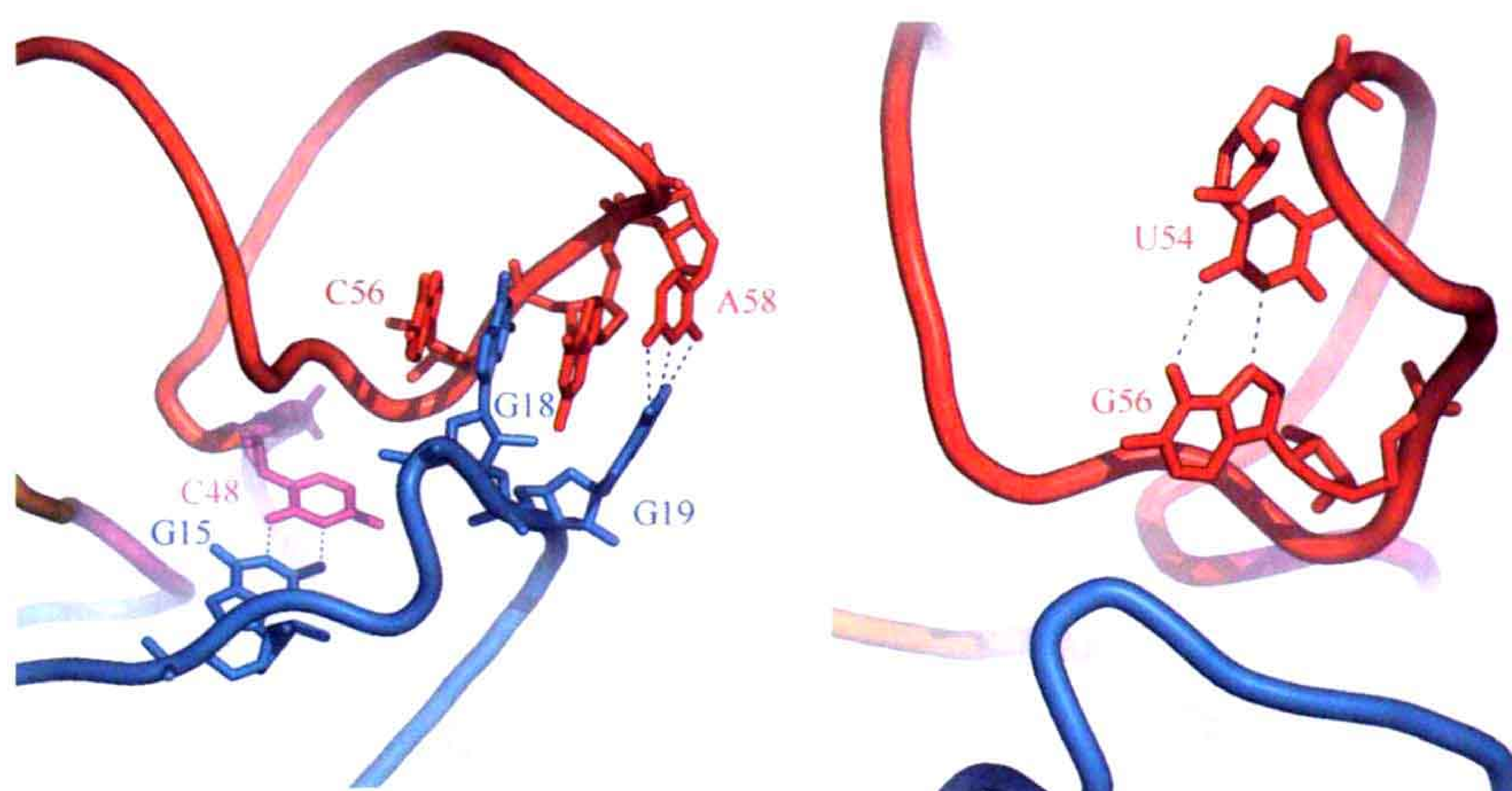


图 3.54　D- 环和 T- 环之间的相互作用。T- 臂上的碱基与 D- 臂上插入的碱基形成了严格的堆积。这里的颜色设置与图 3.53 相同。

T- 环环绕着受体臂，为了稳定这样的结构，一个反 Hoogsteen 碱基配对在环区出现，在 T-54 和 mA-58 位置之间（图 3.54）。突变研究显示，当把这个位置改成可能的等排列（isosteric）UA 碱基对时，它也同样可以是 G-54：A-58 或者 G-54：G58，能稳定三维结构。对此的解释是，这些嘌呤 - 嘌呤碱基对很好地模拟了反 Hoogsteen 碱基对 UA，可因此取代功能 tRNA 中的 T- 环。然而，另一个突变却不能重现结构 54:58 碱基对的原貌，也无法形成稳定的三维结构。

在 tRNA 的骨架上含有 4 个非常大的急转弯，一个是在 D- 环上，一个是在畸变环（variable loop）上，一个是在反密码子环上，最后一个是在 T- 环上。后面的两个是经典的 U- 转弯模体结构，具有稳定的氢键特征，在保守的鸟苷残基和下面链的磷酸骨架之间。

在 tRNA^{Phe} 反密码子环上，核苷酸 34、35 和 36 构成了反密码子三联码 GAA。33~35 位置形成了一个 U 转向，尿苷环内的 N3 和 A-36 上的磷酸氧形成了一个氢键，U-33 核糖上的 O2′ 羟基与 A-35 中的 N7 形成了一个氢键（图 3.55a）。这一 U- 转向导致了反密码子三联码中碱基直接暴露在溶液中，可以与对应的密码子很容易地相互作用。

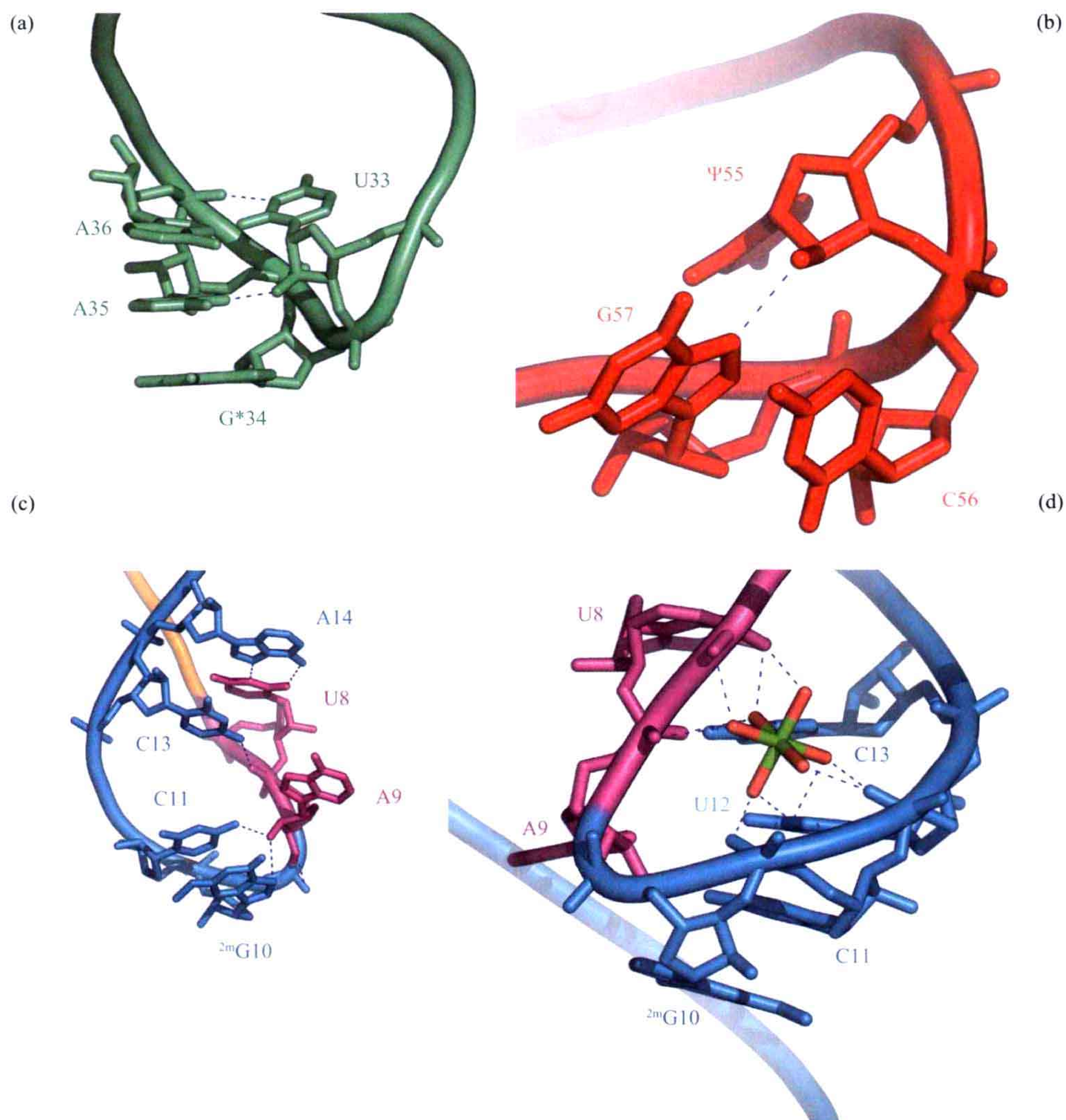


图 3.55 tRNA 结构中的急转弯。(a) 反密码子环 U- 转向 33~36。这一转向是通过 U33 的氢键来稳定的。反密码子碱基延伸到了可见的位置。（b）T- 环的 U- 转向。（c）D- 环前的急转弯。（d）镁离子通过 9~11U- 转向稳定［在与（c）相反的方向观察］。

U- 转向模体也在 tRNA 的 T 环中被找到。残基 55~57 ΨCG 形成了一个 U- 转向，骨架的急转弯由 Ψ-55 的 N3 和核苷酸 58 的磷酸基团之间的氢键来实现（图 3.55b）。在这两个 U- 转向中，尿苷碱基都处在环的内侧，与骨架进行相互作用。tRNA 中骨架最明显的急转弯出现在畸变环中，只出现在了核苷酸 C-48 和 m^{5}C-49 之间。分子两个部分的相互作用有效地使这两个核苷酸的碱基处在了相反的方向上，在主链上形成了

一个非常大的急转弯（图 3.55c）。这一急转弯通过 m^5C-49 与 U-7 之间的堆积相互作用使得在 9~11 位置的转向更加稳定了。在连接区的 9~11 转向导致了 D- 环的出现，这是非常重要的，因为这允许主链从接受臂上折回它的相反方向，折叠成 D- 环的一个叶子在结构骨干的反面(图 3.55c)。由于和相邻核苷酸 8、12 及 13 磷酸基团的配位作用，9-11 转向在镁离子的帮助下形成它的构象(图 3.55d)。最后，9~11 转向通过 D- 臂上不变的 U-8 和 D 环上不变的 A-14 之间的反 Hoogsteen 配对来稳定结构。

A-9 包含在另一种类型的三级相互作用中：它被插入碱基 G：45 和 m^7G：46 之间。为了给这两个碱基提供空间，骨架通过在 m^7G-46 上的 C2′ -endo 糖皱褶来伸展。

这是非常清楚的，U- 转向和 tRNA 结构的转向，在整齐的螺旋结构区域外面，含有丰富的空间相互作用，是形成分子正确折叠形式的重要因素。

tRNA 结构的中心也显示了一些碱基三倍体，它们涉及稳定分子紧密包装的核心。它们处在结构非常核心的位置，即在分子“L”形结构的转角位置。与此有关的一些细节见表 3.3 和图 3.56。

表 3.3　tRNA 中碱基三聚体的详细资料

三聚体	描　述
m^2G10:C25/G45	G45 与 D- 臂中的第一个 Watson-Crick 碱基对的大沟形成了 Sugar/Hoogsteen 型的相互作用
U12:A23/A9	U12:A23 大沟中的 A-9 氢键，与 A23 形成了一个反 Hoogsteen 碱基对。这稳定了碱基 9 和 10 之间的急转弯
C13:G22/m^7G46	畸变环（variable loop）上的 m^7G46 与 D- 臂的 G22 形成了一个 Watson-Crick/Hoogsteen 型的碱基配对相互作用，它是与 C13 碱基配对的。这使得畸变环转向了 D- 臂

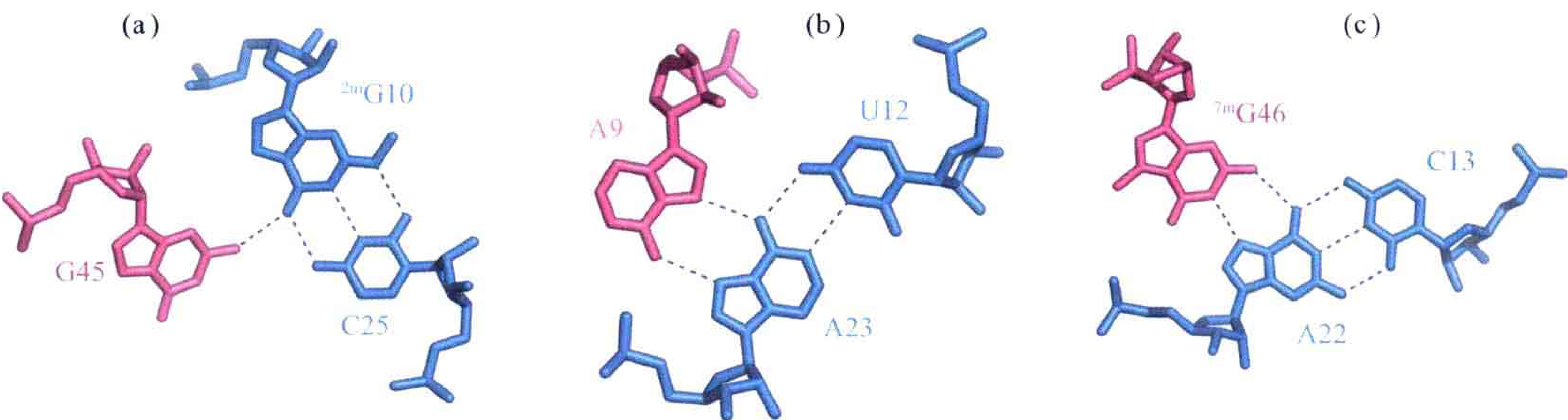


图 3.56　tRNA 的碱基三聚体。（a）m^2G10：C25/G45；(b)U12:A23/A9；(c)C13:G22/m^7G4C13:G22/m^7G46。更多细节见表 3.3。

聚胺类物质，如精胺和亚精胺在 tRNA 三级构型的稳定性中也起到了非常重要的作用，精胺经常会用在 tRNA 或者其他 RNA 的结晶中。在 tRNA 中，有两个亚精胺结合位点：第一个是在深沟处，由 T- 臂和受体臂形成；D- 臂和反密码子臂形成了第二个位点。同时，水分子和二价阳离子也是 tRNA 结构的重要部分。一些水分子间接地与碱基接触，实际上是“延伸”了碱基形成氢键可能到达的位置。六水合镁离子和三水合锰离子也在结构中被找到。

最近的关于 tRNA 结合到延长因子 Tu（EF-Tu）上的研究表明，不同种类 tRNA 躯干的结合亲和力的变化是有意义的（图 3.57，也可见 8.5.1 节）。这些变化其实是由结合到 tRNA 上蛋白质侧链的不同亲和力来平衡的。EF-Tu：tRNA 复合物的晶体数据揭示了与蛋白质有关的大部分信息，包括RNA的骨架，但是在Glu-390和T-臂上G-51：C-63碱基对上的 N2 基团之间只有一个氢键被观察到。*T. thermophilus* EF-Tu 的 Glu390Ala 突变使得那些在 51 或者 63 位置上、含有一个鸟嘌呤的 tRNA 出现了有选择的结合位点。在 tRNA 方面，51:63 碱基对的突变看上去是调节了对 EF-TU 的结合亲和力。与大肠杆菌 tRNA 序列的比较表明，这一专一的机制在细菌中是保守的，这似乎是一个重要的特异性决定因素。

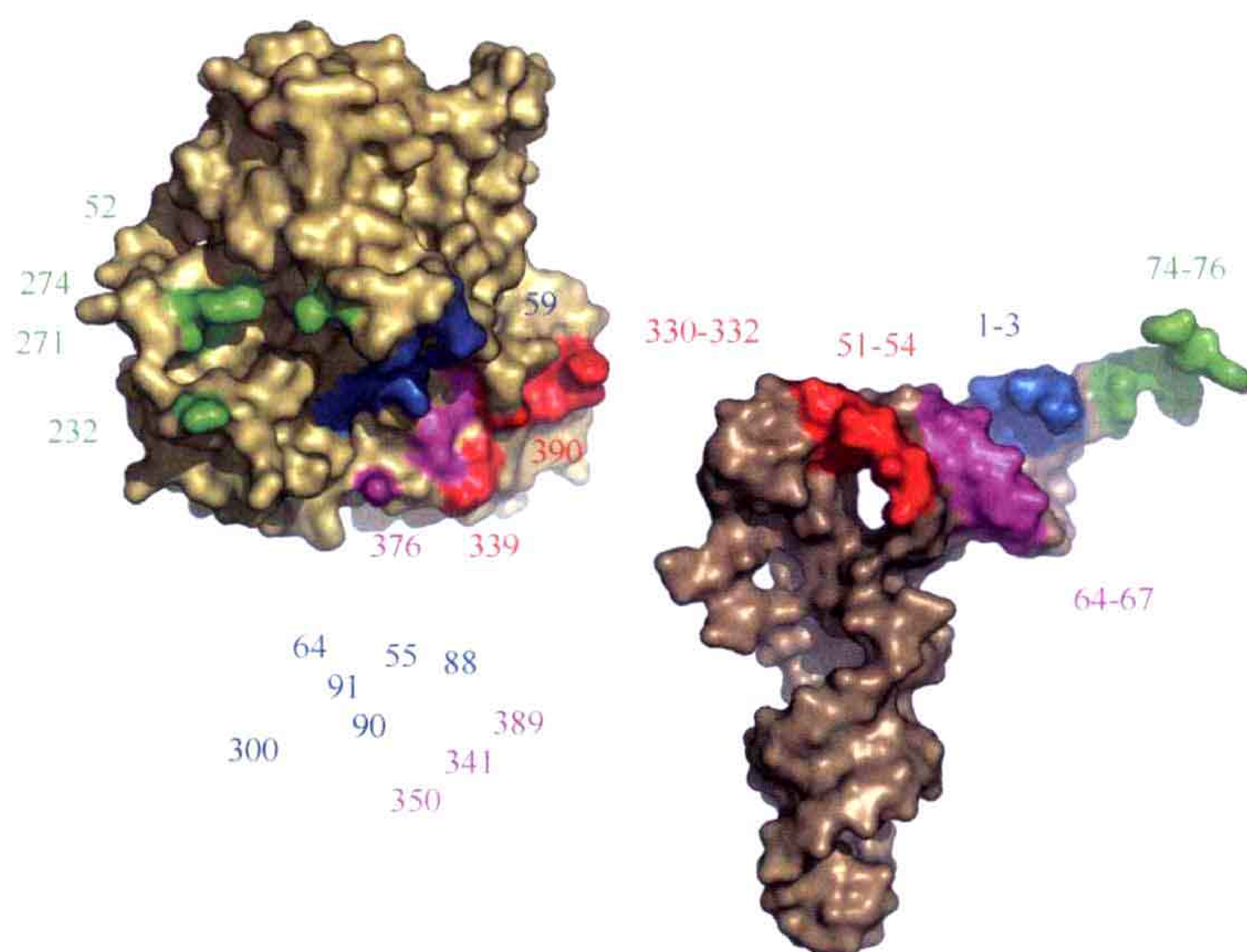


图 3.57　EF-Tu 可以通过骨干与 tRNA 的作用结合到所有胺酰化的 tRNA 上。在 Glu-390 和 G-51：C-63 碱基对之间还出现了一个氢键。右边的图片显示了 tRNA 和 EF-Tu 之间的接触区域。

3.4.11　RNA 具有催化的功能——核糖酶（ribozyme）

第一个核糖酶被 T. R. Cech 和他的团队在 20 世纪 80 年代早期发现，在这之前人们一直相信，只有蛋白质可以催化酶学反应。从那以后，大量关于 RNA 可以催化化学反应的例子被发掘，支持了 Woese、Crick 和 Orgel 提出的“RNA 世界”假说。

已知的具有催化能力的 RNA 包括核糖核酸酶 P（在 tRNA 加工中具有活性）和自分裂核糖酶，如锤头、发夹、HDV 核糖酶和 VS 核糖酶。

mRNA 的组 Ⅰ 和组 Ⅱ 内含子形成了另一个类别的催化 RNA。组 Ⅰ 内含子是一个自剪切内含子，需要 GTP 来帮助其剪切。它包含一个活性位点，可以使它从一个前体 mRNA 切断自身后，再把相邻的外显子相互连接起来。这一自剪切反应可以形成一个完整的 mRNA、tRNA 或者 rRNA。组 Ⅱ 内含子主要在特定的真核细胞细胞器 rRNA、tRNA、mRNA 中，以及细菌的 mRNA 中被找到。核糖体中心的肽转移酶也主要由

rRNA 组成。L1 连接酶是一个人造 RNA 分子，它可以催化 5′ - 三磷酸核苷酸连接到一个 RNA 链的 3′ 端。它可以支持这样的假说，即在生命出现早期，RNA 分子存在于一个“RNA 世界”中，在这个世界中核糖酶承担了自我复制的任务。

3.4.11.1　锤头核糖酶

锤头核糖酶阐明了RNA三级结构的要点。它催化一个酯基转移翻译，在这一反应中，核苷酸 17 和 1.1 之间的 3′，5′ 磷酸二酯键被切断（图 3.58）。锤头核糖酶的第一个结构在 1994 年被解析，其是一个 RNA-DNA 杂合子，底物链被 DNA 链替代。这种杂合的核糖酶不具有催化活性，因为脱氧核苷酸在 17 位缺少了 2′ - 羟基基团，这对于剪切反应来说是非常重要的。如果核苷酸 17 被单个核糖核苷酸取代，则催化活性又恢复。

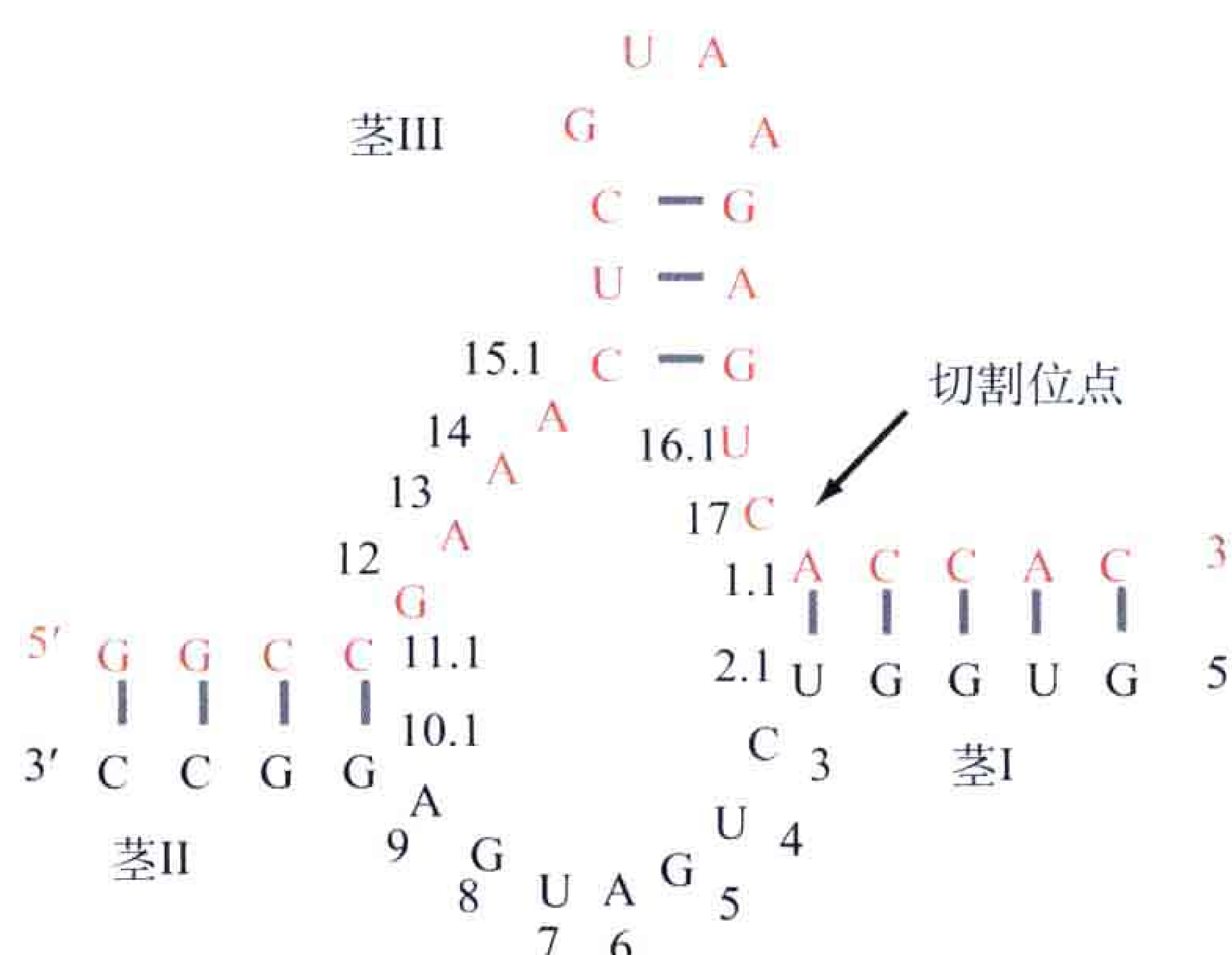


图 3.58　最小的锤头核糖酶的二级结构。底物链被标记为红色。在三连接中心环的保守残基被标记为黑体。关于锤头核苷酸的残基编号也已给出。

另一个锤头核糖酶结构是一个全 RNA 结构，在剪切位点胞嘧啶 17 的 2′ - 羟基被一个 2′ - 甲基基团取代，同时底物 RNA 保持不变。锤头核糖酶是一个最小的、可以在体外构建的结构，它包含一个 16 个核苷酸的酶链和一个 25 个核苷酸的底物链。

除增加了一些氢键相互作用以外，这一结构确实与之前被确定的 DNA 杂合体核糖酶很相似。这一结果被认为是一个核糖酶晶体结构与没有修改的核糖体的真实溶液结构密切相关的依据。

尽管如此，最小限度锤头的结构和它的催化活性之间的关系还是引来了越来越多的争论。由于不可能简单地使得晶体结构和实验相符合，从而出现了两个关于在自然条件下催化活性需要怎样的构象改变的相互矛盾的假说。一个假说认为需要阳离子，如镁离子的参与；另一个则认为需要酸碱催化作用。

2006 年，从人类寄生虫 *Schistosoma mansoni* 中得到的全长核糖体的晶体结构被解析，这进一步提供了催化机理的信息。三茎区域（three stem region）排列成几乎很完美的同轴螺旋。茎 I 部分形成了这一复合螺旋的侧枝，其结构就像冰岛语字母 thorn (þ) 的形状（图 3.59）。基于连接区域的长度，核糖酶的三茎连接属于 A 家族。茎 I 含有一个

凸起，通过氢键和堆积作用与茎Ⅱ的环相互作用。与最小的核糖酶结构相比的话，这一空间相互作用诱导了一个催化活性位点的有意义的改组。在骨干区域出现一个急转弯，其在 C-17 处向右绕过易断裂的键。

连接的中心环上的碱基与活性位点密切相关，虽然从二级结构图上无法看到（图 3.59）。残基 C-3、U-4 和 G-5 形成了一个很大的 U- 转向，这些碱基朝向 C-17 暴露在外侧的核糖来帮助核苷酸就位。G-12 内环的 N1 与 C-17 的 2′ - 羟基基团之间处在氢键键长范围之内，这是反应的亲核进攻。G-8 的 2′ - 羟基基团与易断裂磷酸上的 5′ 氧形成了一个氢键，这说明 G-8 可能参与了离去基团的稳定作用。

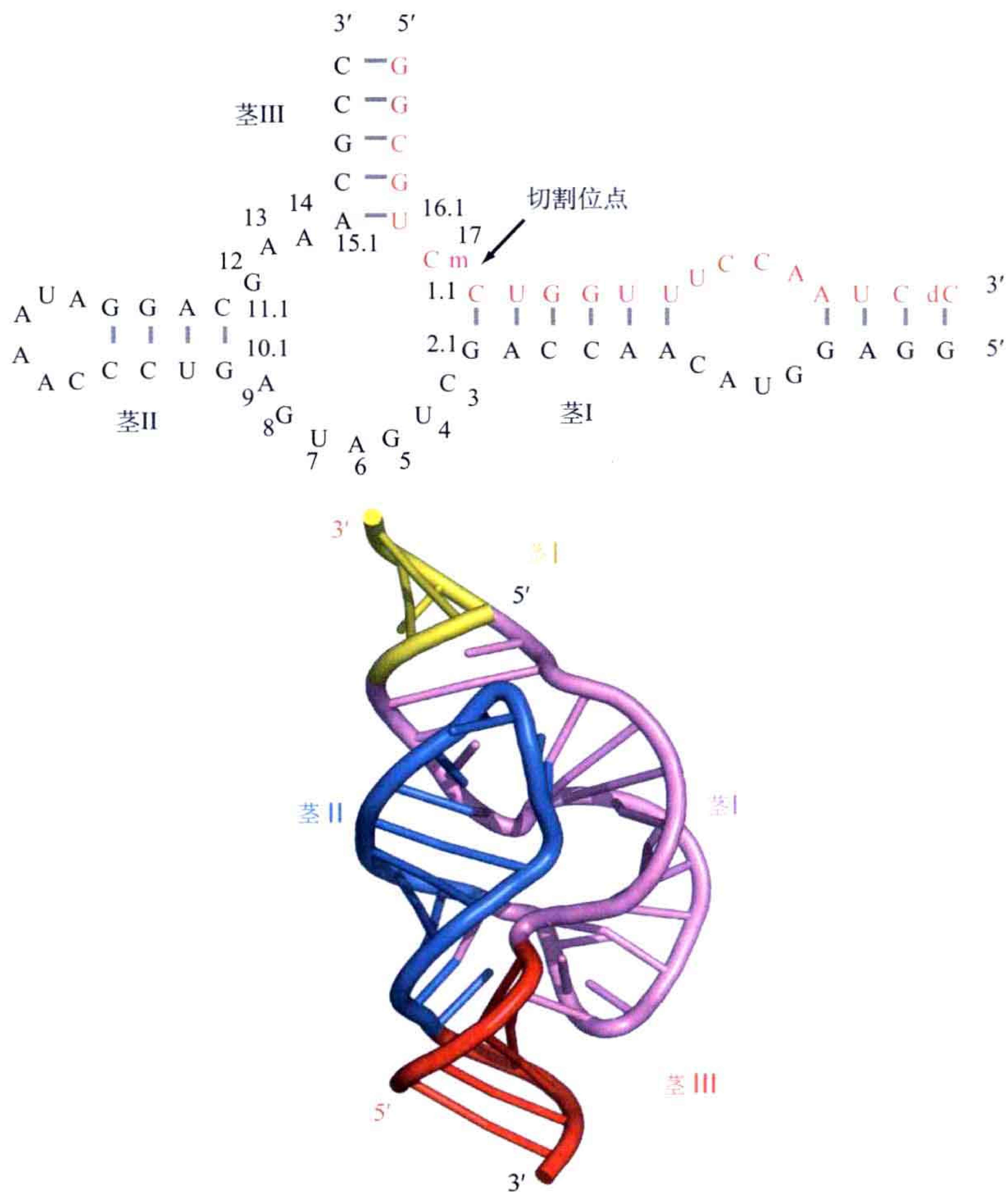


图 3.59 上：全长核糖酶的二级结构图（PDB：2GOZ）。红色核苷酸是底物链，黑体字母表示保守序列模体。在底物 3′ 端的小写字母“d”表示一个脱氧核糖核苷酸，用来提高合成产率。活性位点被标记为 Cm，表示它是甲氧基化的。下：全长核糖体的图示。注意三茎形成了一个同轴螺旋，分别用红色、蓝色和黄色表示。

锤头核糖酶的催化机理见图 3.60。在反应的开始，C-17 的 2′ -OH 必须被活化，这是 G-12 的任务，其内环中的 N1 就在附近。然而，在它正常的质子化状态时，鸟嘌呤

在 N1 位有一个氢原子，但是它的烯醇互变异构体没有。假设 N1 可以活化 C-17 的 2′ -羟基，则这一基团可以形成对连接在相同碱基 3′ - 羟基上的易离去磷酸的亲核进攻。这一反应形成了一个在核苷酸 C-17 和核苷酸 1.1 上自由 5′ - 羟基之间的环状 2′，3′ -磷酸二酯键。这一催化模型是一个共线（in-line）的机制，这使得亲核进攻基团、磷酸盐原子和离去基团在一条线上。反应中间产物的外形是一个三角双锥，它的顶点位置分别是离去基团 (5′ -OH) 和进攻基团（2′ -OH）。

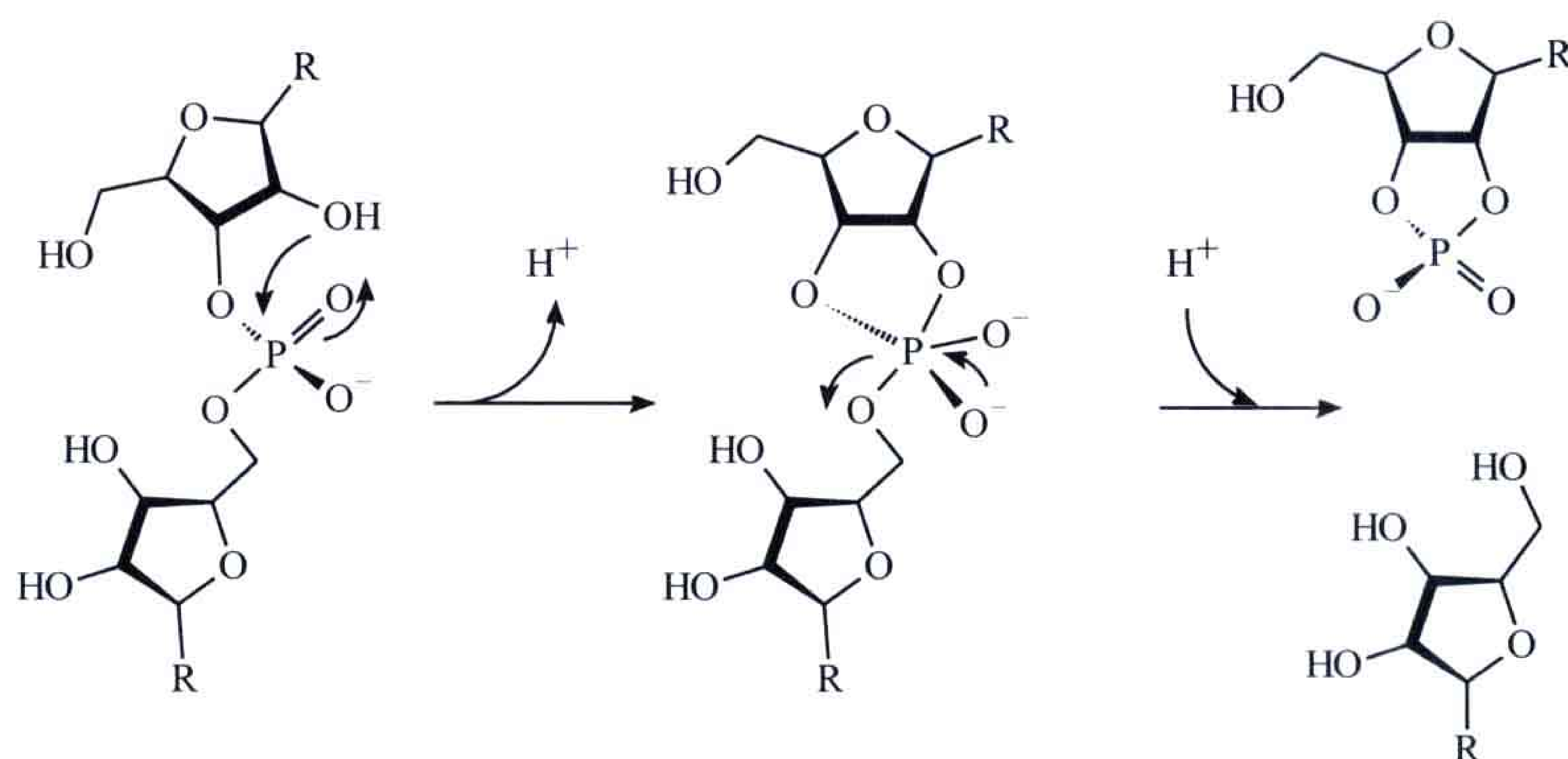


图 3.60　建议的锤头核糖酶的催化机制。

图 3.60 显示了在活性位点周围重要碱基的位置。在晶体结构中，C-17 的 2′ -OH 基团被甲氧基化了，因此反应不能进行。核苷酸 A-9 用它的 N1 基团稳定地朝向 G-12，很容易被活化。氢键和堆积相互作用使得 RNA 在可剪切的磷酸骨架上形成了一个急转弯。这样一个骨架转弯在其他一些结构中也被观察到过，如核酸酶。甚至可以想象成骨架刚性得（“snapping”）像个棍子。图 3.61 也同样显示了一个假定的反应过渡态的化学结构图，碱基 G-12 和 G-8 与过渡态的稳定有关。

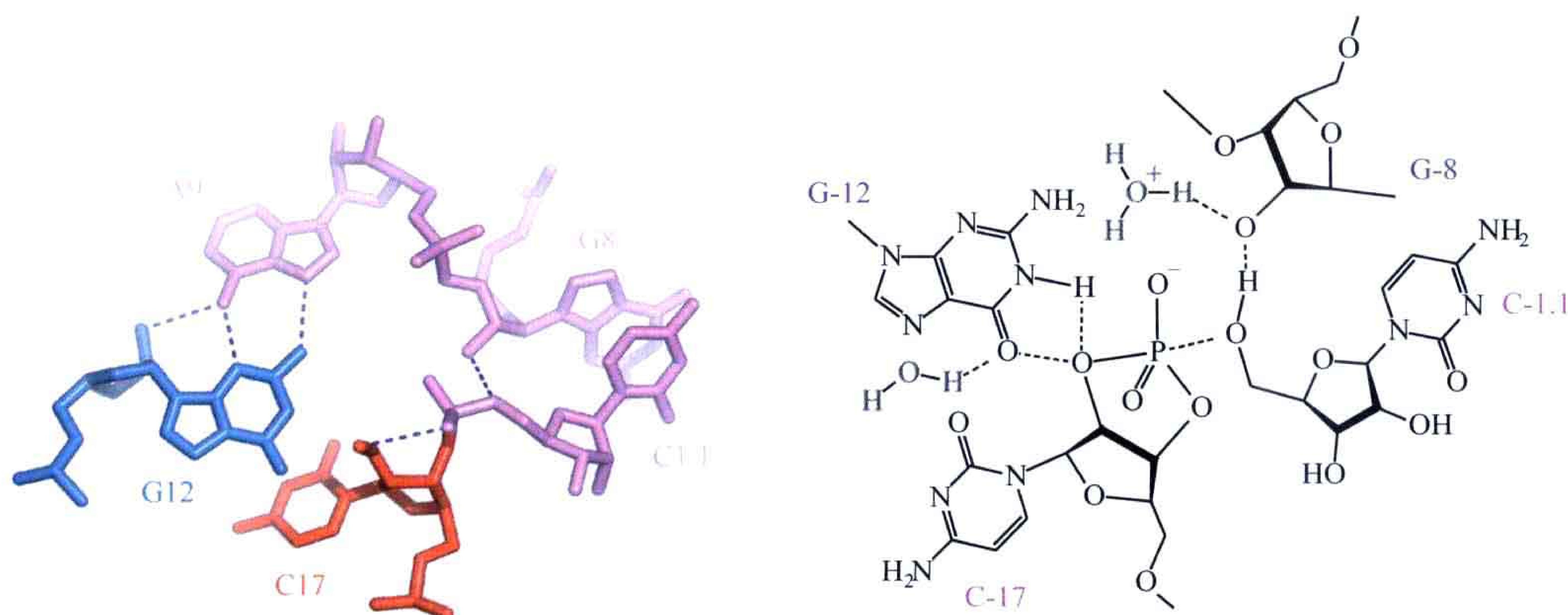


图 3.61　左：全长核糖酶活性位点的结构细节；右：假设的过渡态结构，G-12 正在开始亲核进攻，而 G-8 正在稳定离去极端 C-1.1。

最小的核糖酶的结构没有错误。在一个可信的分辨率下对它的晶体学分析也完全

正确。但是假设这一最小的核糖酶结构与全长核糖酶相同却出乎意料地出现了错误。30 年的蛋白质晶体学实验告诉我们，当蛋白质亚基形成一个大结构时，其结构很少会改变。RNA 结构更加复杂，也很难被预测，最小结构和全长结构之间的差别虽然是细微的，但却是很重要的。当三茎连接结构的全折叠接近完全相同时，附加的在全长核糖酶中的空间相互作用也会导致一个有意义的构象变化，在活性位点区域，生物化学实验证明对于反应来说有一个碱基是很重要的，而现在这个碱基出现在最小核糖酶接近活性位点大概 20Å 的位置。

结构和生化实验的不一致无论何时都需要认真对待，它们不一定表示结构是不正确的，但是可能有些事情并不是看到那样，有可能一个非常有趣的故事正隐藏在数据中。

3.4.11.2 P4-P6

Tetrahymena 组 I 内含子 mRNA[1GID] 中含有 160 个核苷酸的 P4-P6 域的分辨率为 2.8Å 的晶体结构提供了一些 RNA 折叠模体更加细致的信息，也给出了关于前面章节中讨论的 RNA 结构基本原理的一些例证。

组 I 的自剪切内含子催化它们从一个前体 mRNA 的自我切除。组 I 内含子可以被归类为核糖酶。*Tetrahymena* 组 I 内含子的催化核心包含两个主要的域，催化位点就位于它们之间。因此，P4-P6 结构域只包含一半的活性位点，所以它自身没有催化活性。

图 3.62 显示了 P4-P6 的结构，包括碱基对片段 P4、P5 和 P6，以及链接区域 J3/4、J4/5 和 J5/6。P4-P6 也会含有一个延伸的部分，即 P5abc(P5a、P5b 和 P5c)，其只在组 I 内含子的一个亚类中被找到，但是它一旦出现，对于催化活性是非常重要的。

在 P4-P6 结构中，有两个螺旋区域，肩并肩排列，占主导地位（图 3.63）。通过同轴堆积，螺旋 P6b、 P6a、 P6、P4 和 P3 在分子的一侧形成了一个直的圆柱体（保守核心），而延伸的部分 P5abc 形成了一个香蕉状的支持部分。

两个重要的空间相互作用形成了一个在延伸片段和保守核心之间的界面。其中之一是在 P4 螺旋的浅沟和富含 A 的凸起的碱基之间的相互作用。另一种相互作用来自 P5abc 片段尖端的 GAAA 四环和保守核心的四环受体，其位于螺旋 P6a 和 P6b 之间的内环上（J6a/6b）。

在晶体结构中，富含 A 的凸起形成了一个转向，其中的 A 碱基都被倒置于外面。凸起的 4 个 A 碱基与堆积相互作用和氢键的形成都有关，是分子中两个螺旋堆积的有效桥梁。为了稳定富含 A 的凸起的紧急转向，两个镁离子直接与环上的磷酸氧形成了配位键。在凸起中，最前面的两个 A 碱基结合到了螺旋 P4 的深沟上，而最后两个碱基则结合到了延伸片段三茎连接的口袋中。

对于富含 A 的凸起在稳定 P4-P6 结构域三维结构重要性方面的观察，很好地与定点突变和化学保护实验相吻合，这也强调了整个结构对于凸起上的突变非常敏感。

GAAA 四环构型很容易让人想起在 NMR 研究中观察到的，以及在锤头核糖酶中相似的结构，可以确定 GNRA 折叠确实是一个刚性实体。四环与结构域保守核心中的含有 11 个核苷酸的受体片段相互作用。这三个突出的腺苷酸是处于 *anti* 构象的，它们进

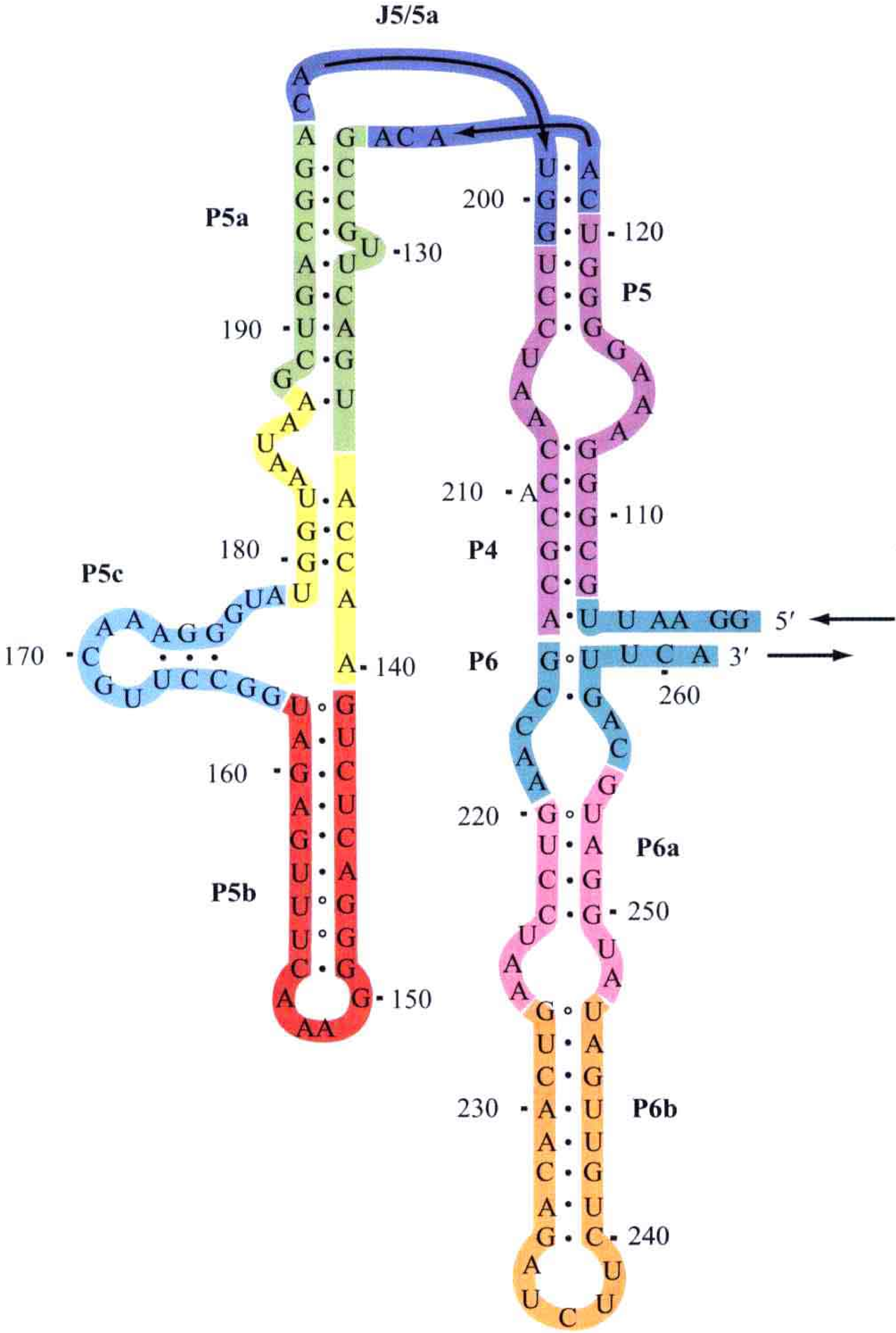


图 3.62 组 I 内含子 P4-P6 结构域的二级结构图示。

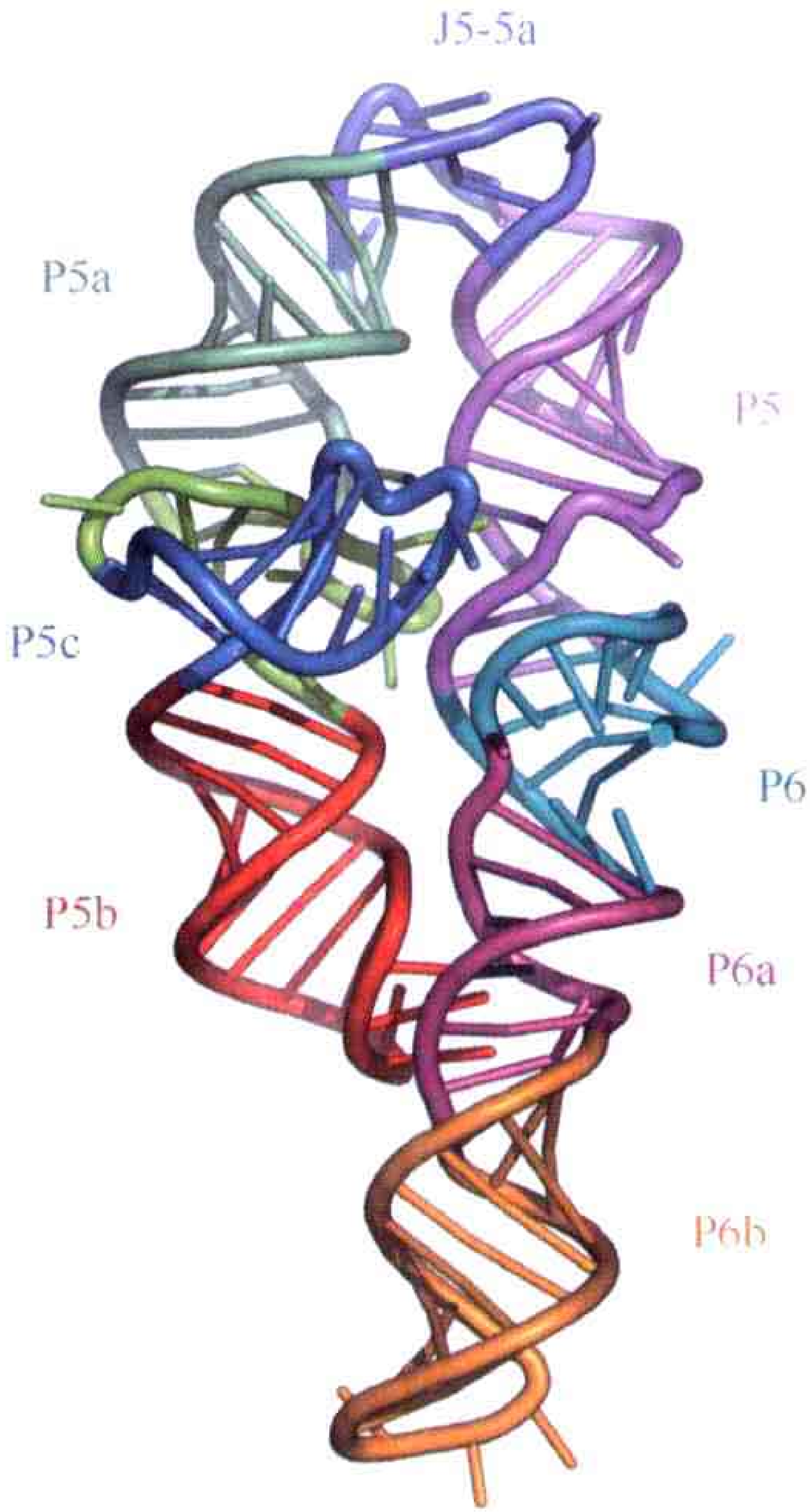


图 3.63 在与图 3.62 展示的二级结构相同方向观察 P4-P6 分子的三级结构。

入 P5b 和 P6b 螺旋的浅沟中，可以与受体片段中的碱基形成堆积相互作用。这一堆积被受体内环上的相邻腺苷酸促进，可以形成一个特殊的三碱基相互作用，这就是被称为 A-A 平台（A-A platform）的结构（图 3.64）。除了堆积，GAAA 环的每个腺嘌呤与受体都形成一个特定的氢键。

富含 A 凸起和 GAAA 环使得 P4-P6 分子圆柱体的两半部分很近地靠在了一起，导致了核糖磷酸骨架之间非常紧密的缠绕。两种基本类型的相互作用在 RNA 结构中很常见，可以稳定这样的排列。

首先，磷酸与毗邻的螺旋的紧密缠绕是由水合镁离子解除的。除此之外，来自结晶缓冲液中的两个钴原子被观察到，其也很适合结构中的口袋。金属离子——特别是镁离子，在大量的 RNA 结构中被发现，其对 RNA 折叠成稳定的三级结构及一些 RNA 酶的催化活性来说是必需的。

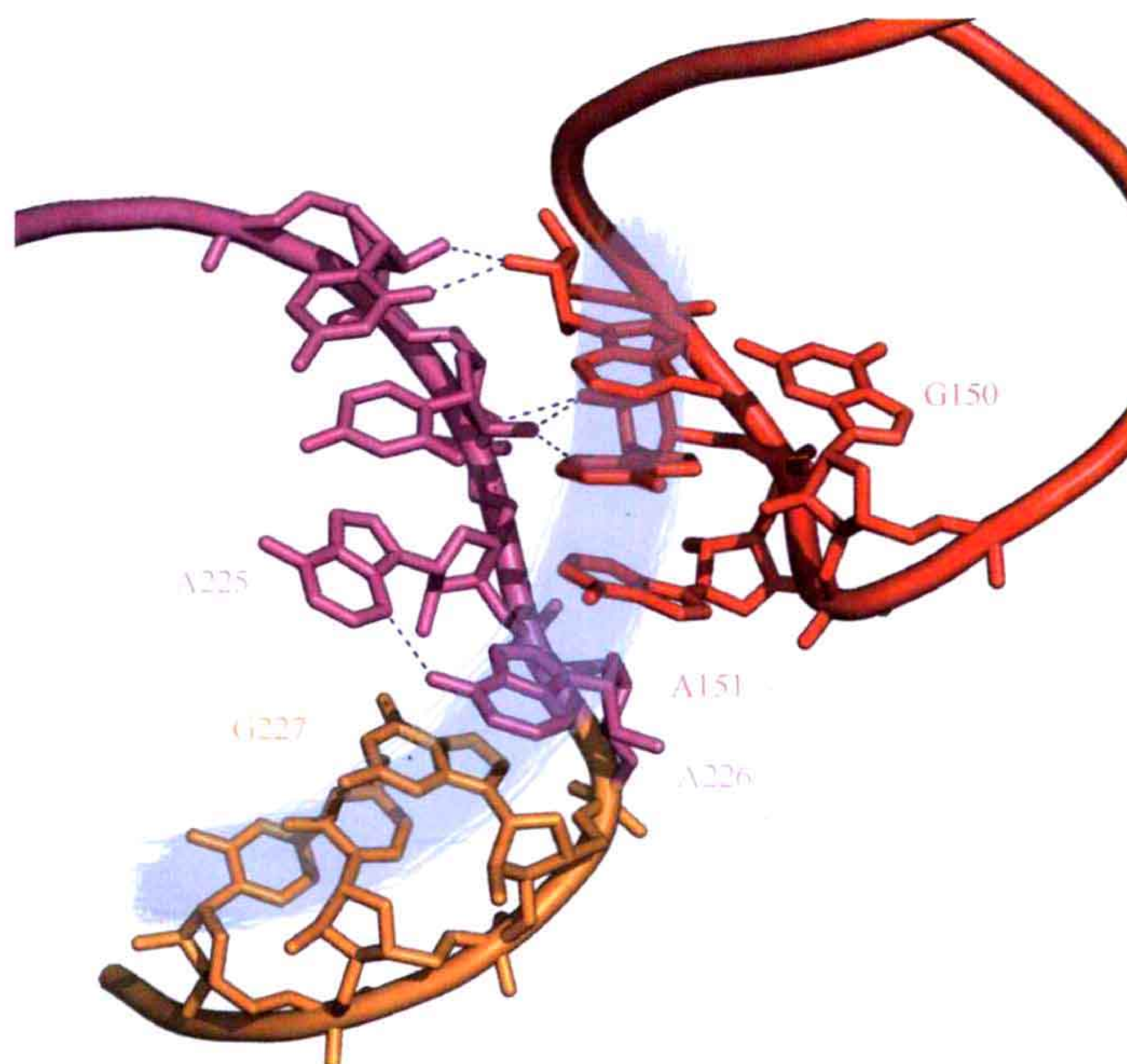


图 3.64　A-A 平台碱基。包括 A225 和 A226，促进了一个外加的疏水堆积，这一相互作用用蓝色表示。

其次，在富含 A 的凸起和 GAAA 四环模体中，来自两个螺旋交错堆积的核糖单元在 2′-羟基基团之间成对形成了氢键。这一结构模体被称为核糖拉链（ribose zipper），它主要在一些大 RNA 结构中被发现，包括 50S 核糖体（图 3.65）。

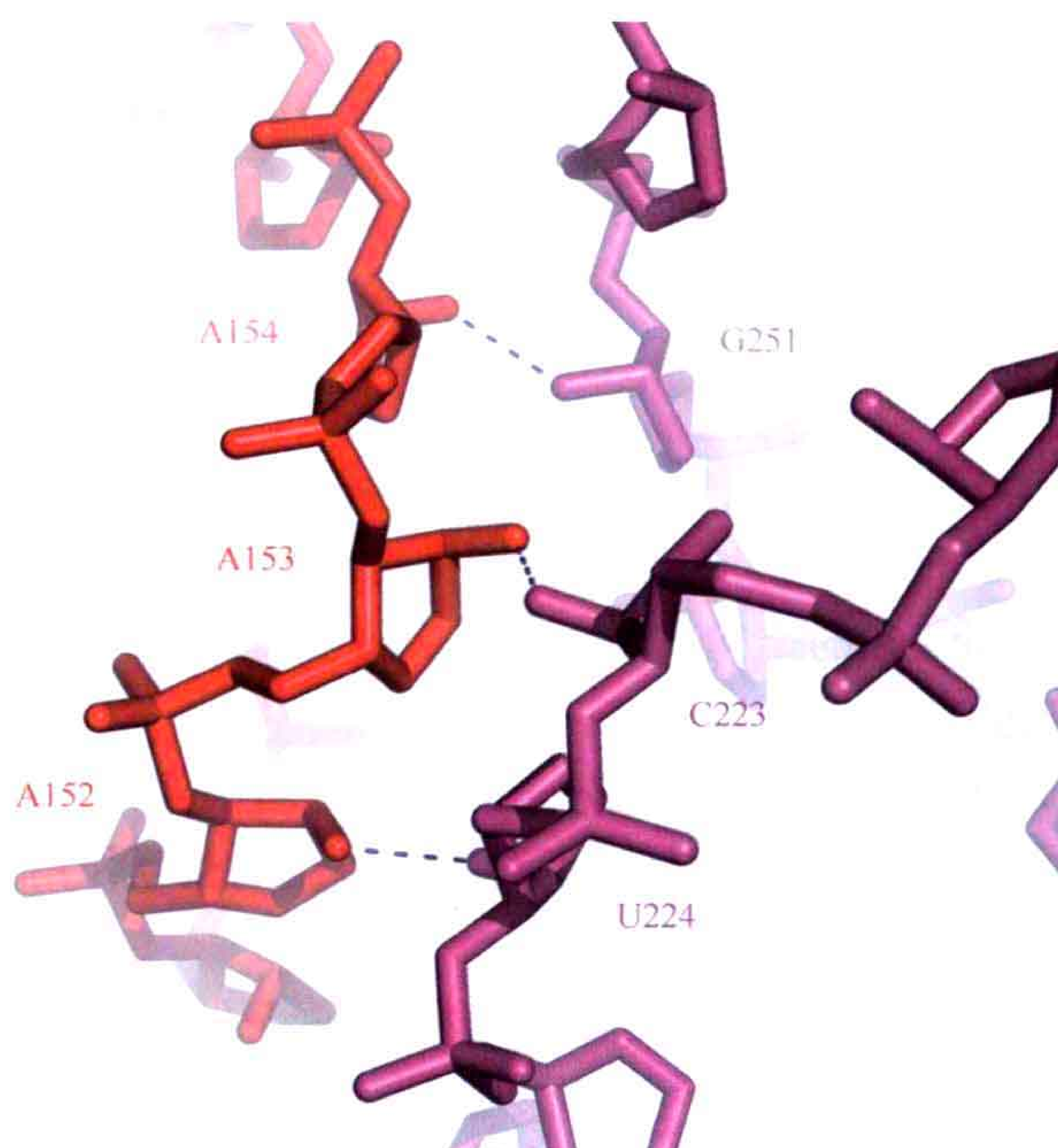


图 3.65　P4-P6 中的核糖拉链模体的特写。

P4-P6 结构域的结构清楚地阐释了 A 碱基在形成空间相互作用甚至是中等大小 RNA 中的重要性，事实上，上面讨论的 5 种相互作用只是 A- 小模体的一些例子。

延伸阅读（3.2 节）

综述文章

Dickerson RE. (1989) Definitions and nomenclature of nucleic acid structure components. Nucleic Acids Res **17:** 1797–1803.

延伸阅读（3.3 节）

原始文献

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature **389:** 251–260.

Watson JD, Crick FHC. (1953) Molecular Structure of Nucleic Acids: A structure for deoxyribose nucleic acid. Nature **171:** 737–738.

延伸阅读（3.4 节）

原始文献

Cate JH et al. (1996) Crystal structure of a Group I ribozyme domain: Principles of RNA packing. Science **273:** 1678–1685.

Leontis NB, Westhof E. (1998) Acommon motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. J Mol Biol **283:** 571–583.

Nissen P et al. (2001) RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. Proc Nat Acad Sci USA **98:** 4899–4903.

综述文章

Leontis NB, Westhof E. (2001) Geometric nomenclature and classification of RNA base pairs. RNA **7:** 499–512.

Hendrix DK, Brenner SE, Holbrook SR. (2005) RNA strutural motifs: Building blocks of a modular biomolecule. Quart Rev Biophys **38:** 221–243.

Lescoute A, Westhof E. (2006) Topology of three-way junctions in folded RNAs. RNA **12:** 83–93.

数据库链接

NDB: http://ndbserver.rutgers.edu

SCOR: http://scor.berkeley.edu

RNA modification database: http://library.med.utah.edu/RNAmods

Non-canonical base pair database: http://prion.bchs.uh.edu/bp_type

Metal binding sites: http://merna.lbl.gov

（金坚石 译　苏晓东 校）

第四章

脂类及膜结构基础

4.1 简介

100 多年以前，查尔斯·欧沃堂（E. Charles Overton）发现非离子或者中性分子穿过植物细胞膜的速率与这些分子在液态脂肪中的溶解度密切相关。他在对蝌蚪的实验中发现，一般麻醉剂的作用程度与其在水和橄榄油之间的分配比率有关，并因此得出膜的主要成分是脂质 (lipid，其定义见下文。在这里简单地说，“脂质”指的是一种由亲水和疏水部分组成的油脂状分子) 的结论。这种活细胞与其所在环境之间的一层仅为 7~10nm 厚的分界线，被称为质膜（plasma membrane）。

这层膜组成的屏障控制着营养成分和其他分子进出细胞。它可以感受激素和其他外界信号，主要由磷酸酯类及嵌入其间的蛋白质组成。1925 年，Gorter 和 Grendel 提出这些膜脂以两个分子厚度的片层状形式存在并将之命名为脂双层。如图 4.1 所示，疏水、非极性的烃链尾部粘在一起形成两分子厚的脂双层，而亲水、极性的头部形成与水溶液之间的界面。其他一些常见的脂质的聚合形式也如图 4.1 所示。

生物体不同膜之间的弯曲程度有很大差异。其中一些如线粒体或者叶绿体的类囊体部分的膜有很大弯折（图 4.2）。这种剧烈弯折的膜与扁平的脂双层的构造单元是不一样的。

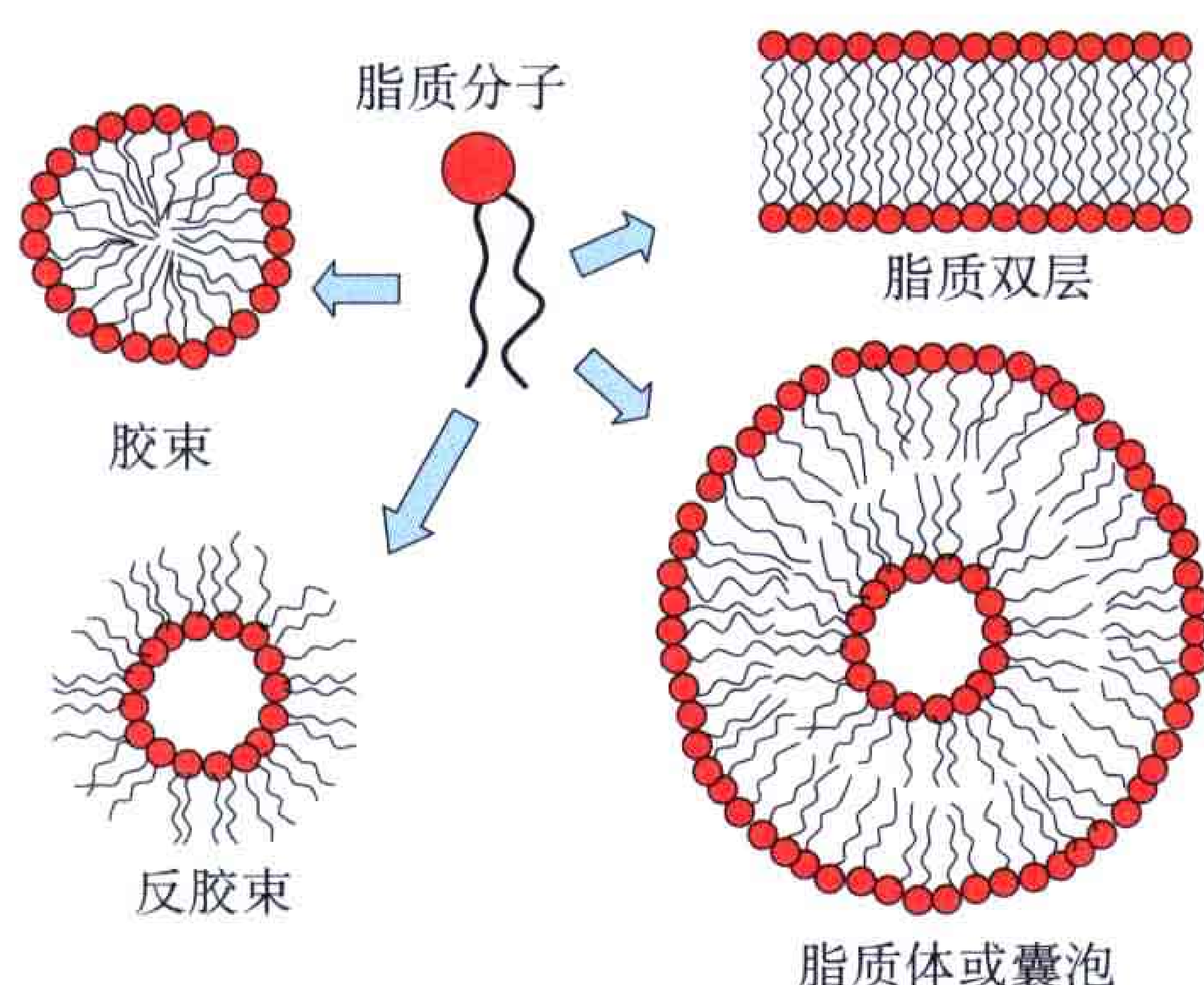


图 4.1　4 种典型脂质聚集结构示意图。脂双层也可以形成闭合的结构，称为脂质囊泡或者脂质体。注意这些图仅仅显示平均的几何结构。实际上，这些结构的变化和动态要复杂得多。

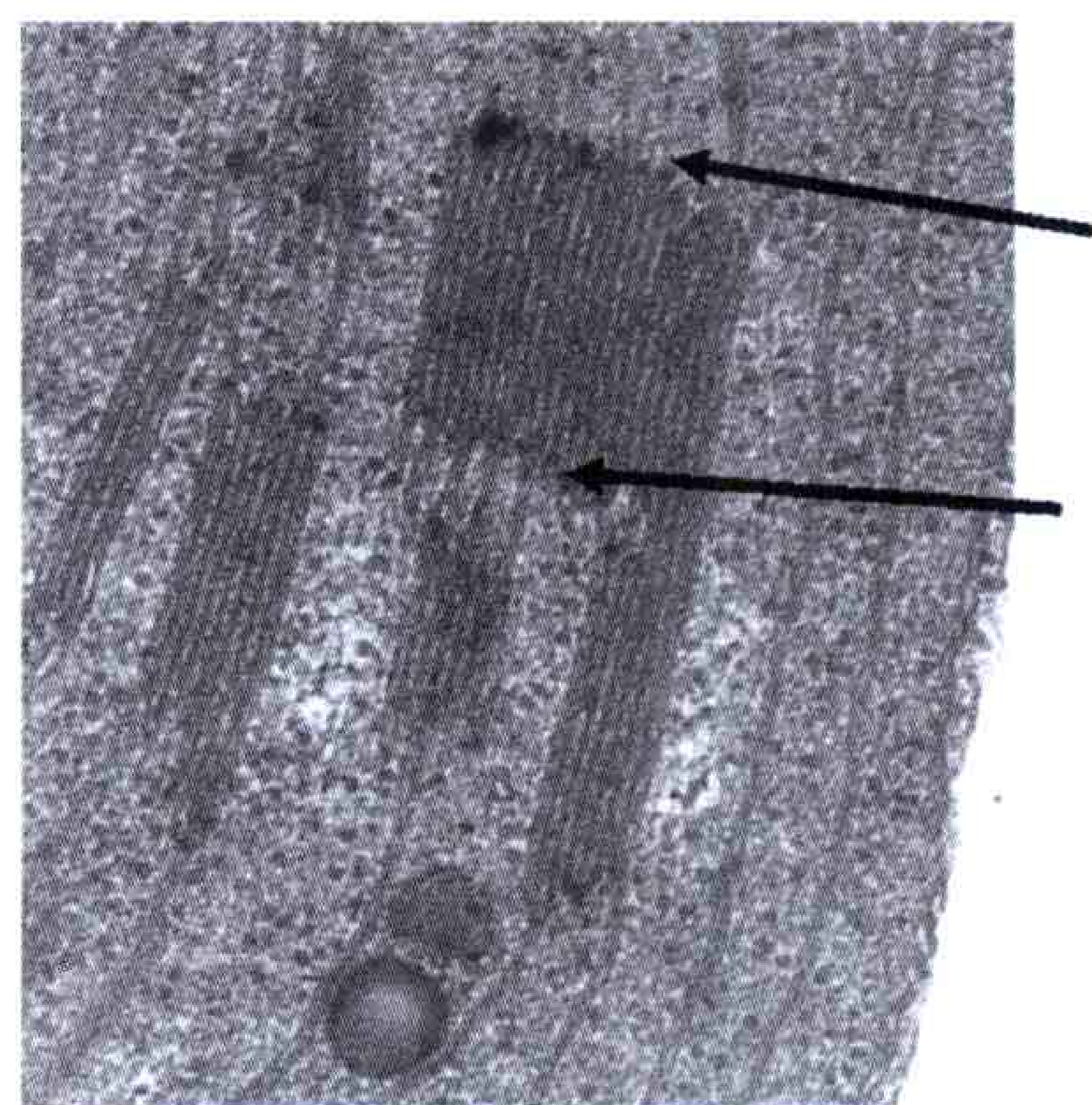


图 4.2　来自线粒体的类囊体膜上在扁平区域的连接处有急剧的弯折（如箭头所示）。电镜照片由 C.Weibull 所摄，P.Å. Albertsson 惠赠。

本章将提供以下几方面的基础知识：

——脂质（为何）会聚集；

——不同的结构形式（何时）形成；

——脂质不同的物理化学性质（如何）被细胞利用。

几乎所有的生物膜都以脂双层的形式存在。最常见的脂质分子是脂肪（fat），它通常是动物及多数植物中用于储藏能量的化合物。一些脂质（由长链脂肪和羟化脂肪酸形成的复杂混合物）可以形成植物的外部表皮，还有一些（如蜡）可以作为羽毛和毛发的保护层。另外，维生素 A、D、K、E 和泛醌，一些激素分子，以及植物色素如叶绿素和类胡萝卜素等都是脂质。

4.2 膜的组成分子

4.2.1 脂质分子的分类及特点

脂质是什么？在相当长的时间内对脂质的定义都是一个难题。磷脂是一种典型的脂质，而胆固醇同样是脂质吗？最近，有文献试图给脂质一个全面的定义（见“延伸阅读”）。按照其化学功能骨干的不同，脂质可以分为以下 5 种，即聚酮类、甘油酯类、鞘脂类、萜醇类及糖酯类。然而，由于一些历史原因及在生物信息学方面的一些优势，人们将属于聚酮类的脂肪酰类、属于甘油酯类的甘油磷酸酯及属于萜醇类的固醇类单独取出作为独立的分类，所以形成了一共 8 种基本脂质种类（表 4.1 和图 4.3）。

表 4.1 脂质分类和举例

分 类	举 例	图 4.3 中的结构
脂肪酸	十六烷酸	(a)
甘油酯	1- 十六酰基 -2-（9Z 十八酰基）-sn- 甘油	(b)
磷酸甘油酯	1- 十六酰基 -2-（9Z 十八酰基）-sn- 甘油 -3- 磷脂酰胆碱	(c)
鞘酯	*N*- 十四酰基 -4- 鞘胺醇	(d)
固醇酯类	5-en-3 β- 胆固醇	(e)
萜醇类	2E-6E-10E 法呢醇	(f)
糖脂类	UDP-3- 氧 -（3R- 羟基十四酰基）- α D-*N*- 乙酰氨基葡萄糖	(g)
聚酮类	黄曲霉素 B_1	(h)

注：Z 代表顺式双键；E 代表反式双键。

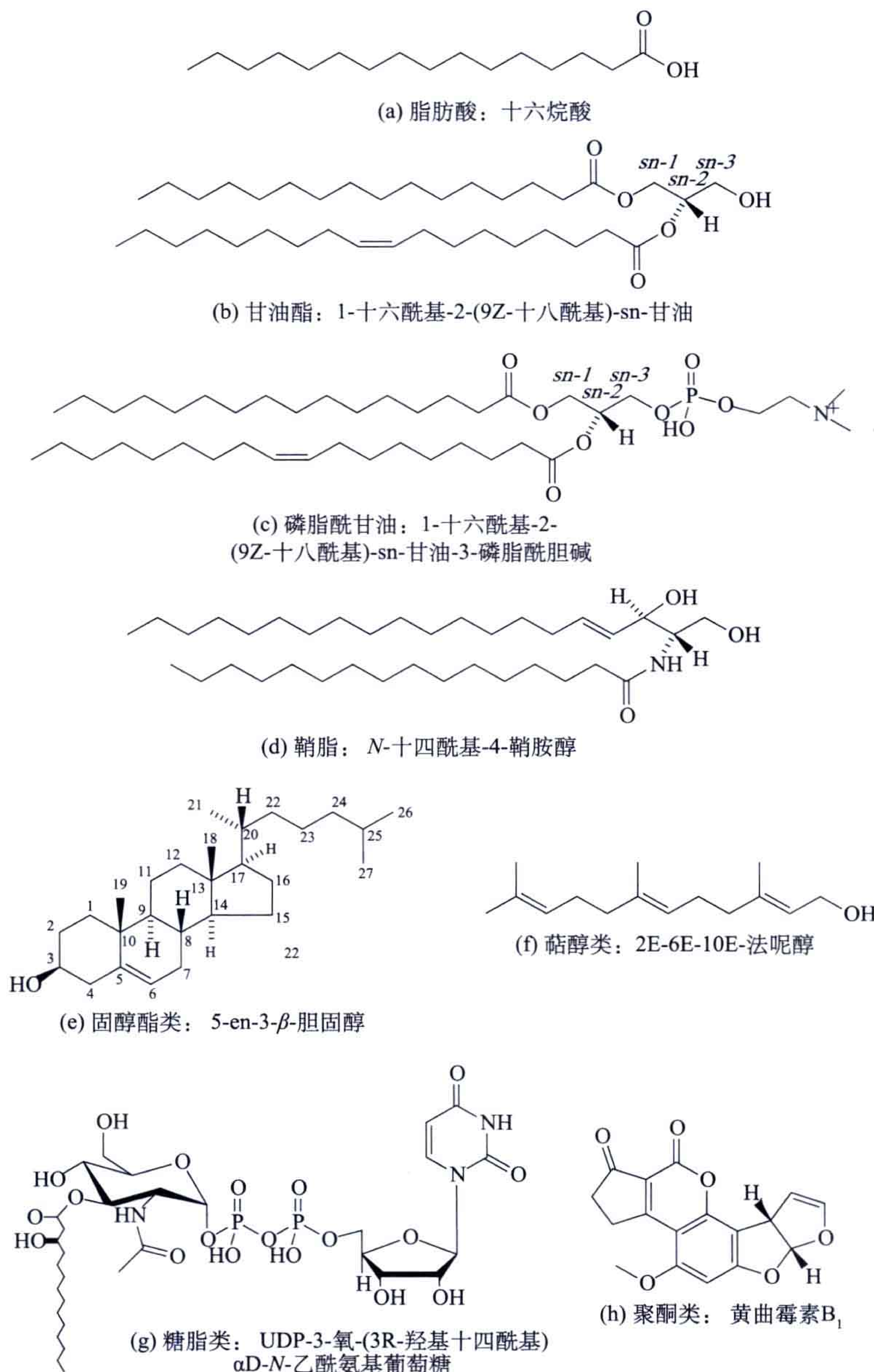


图 4.3　每种脂质的代表结构 [经授权根据 Fahy *et al*. (2005) A comprehensive classification system for lipids. *J Lipid Res* 46: 839-861. Copyright (2005) American Society for Biochemistry and Molecular Biology 改编]。

发展一套系统的脂质分类系统和系统命名体系，首先需要建立清楚的脂质结构绘制方法。有几个网站提供了有用的联机资源（见“延伸阅读”）。这里，我们提供一种简化的脂质结构版本，足以对细胞膜的构造单元进行理解。与蛋白质、多糖和核酸不同，多数脂质不是多聚体，而是由一些被称为“构造单元”的小分子零件连接组合而成的。脂质的“构造单元”包括脂肪酸、甘油、磷酸和糖。脂质由极性区域和非极

性区域构成，从而赋予了其两亲的特征，其在水中倾向于形成膜状的结构或者其他的液晶相结构。

绝大部分膜脂都含有两个连接在甘油骨干上的脂肪酰链。值得注意的是，虽然甘油本身不是手性分子，但它上面的两个羟甲基化基团的位置是不相同的（图 4.4）。图 4.5 展示了某些重要的饱和及不饱和脂肪酸。

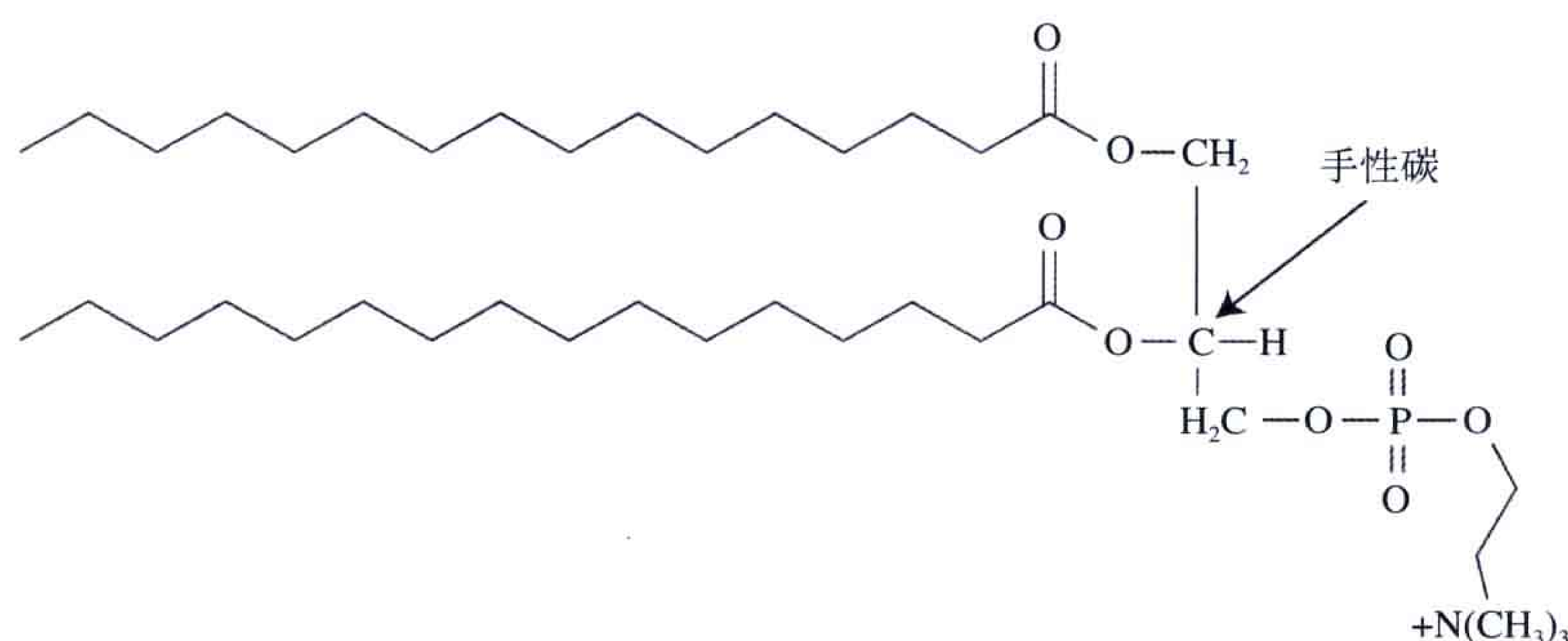


图 4.4　一种二棕榈酰磷酸酯分子的结构，甘油中的手性碳原子如箭头所示。

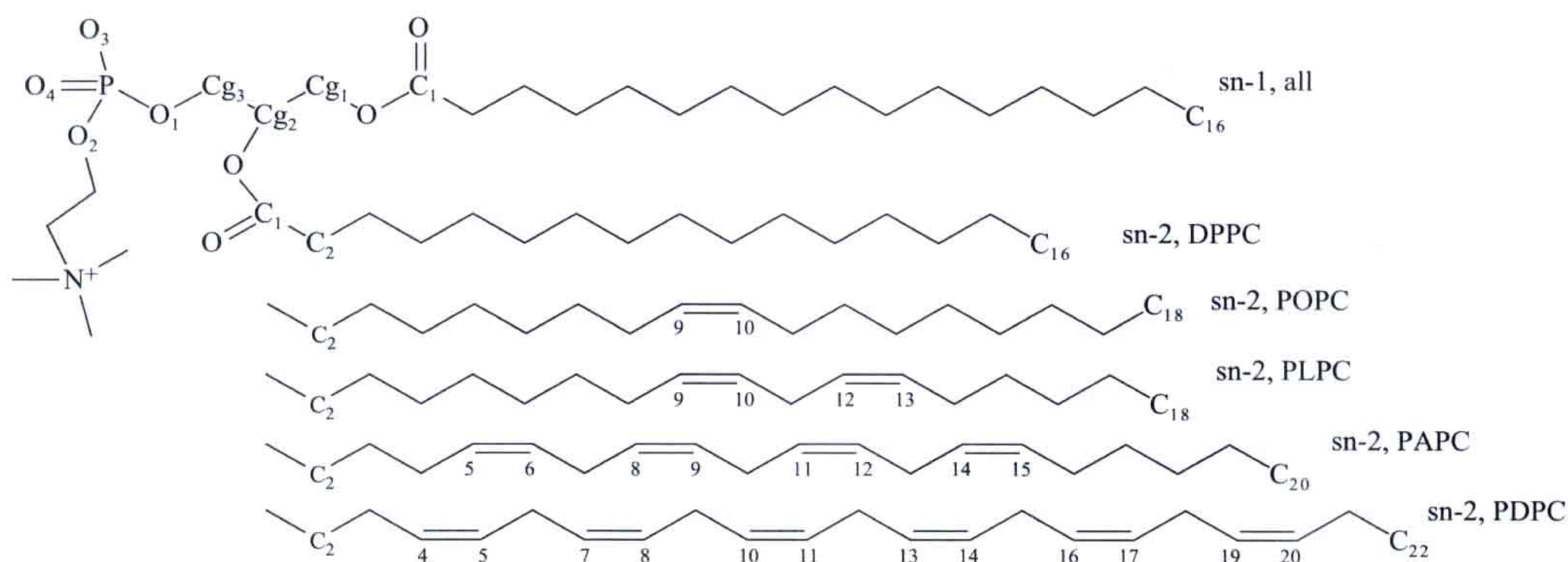


图 4.5　磷脂酰胆碱中常见的脂肪酸酰基链。DPPC 代表二棕榈酰脂酰胆碱；POPC 代表软脂酰油酰胆碱；PLPC 代表软脂酰亚油酰胆碱；PAPC 代表软脂酰花生四烯酰胆碱；PDPC 代表软脂酰二十二碳六烯酰胆碱。

脂肪酸的种类从结构上说几乎有无数种可能存在，但是在特定的物种中主要只有少数几种存在。由于脂肪酸合成途径中每个循环添加一个二碳单位，多数的脂肪酸都含有偶数个碳原子（见 5.4 节）。植物和动物主要含有 C_{16} 饱和棕榈酸，以及 C_{18} 顺式－不饱和油酸。除此之外，动物中还含有大量的 C_{18} 饱和硬脂酸，以及少量的 C_{20}、C_{22}、C_{24} 脂肪酸。一些多不饱和脂肪酸是人类食物中必不可少的。其中，花生四烯酸是被称为前列腺素激素及一系列相关的前列腺素类化合物的合成前体。我们脑中的磷脂分子含有大量的二十二碳六烯酸。细菌通常不含有多不饱和脂肪酸，但是经常含有支链脂肪酸、含环丙烷结构的酸及羟化脂肪酸。

复杂的脂质中各成分连接的方法有多种，通常以甘油作为核心组成单位。因此，一般的脂肪组织中的脂和植物中的油都是三酰甘油或者又称甘油三酯，其甘油骨干上

有三个以酯键连接的脂肪酸。

4.2.1.1　磷脂

作为生物膜的主要成分，磷脂在所有的活细胞中都有重要的作用。磷脂有两个基本种类，即含有甘油的甘油磷脂和含有神经鞘氨醇的鞘磷脂（图 4.6）。

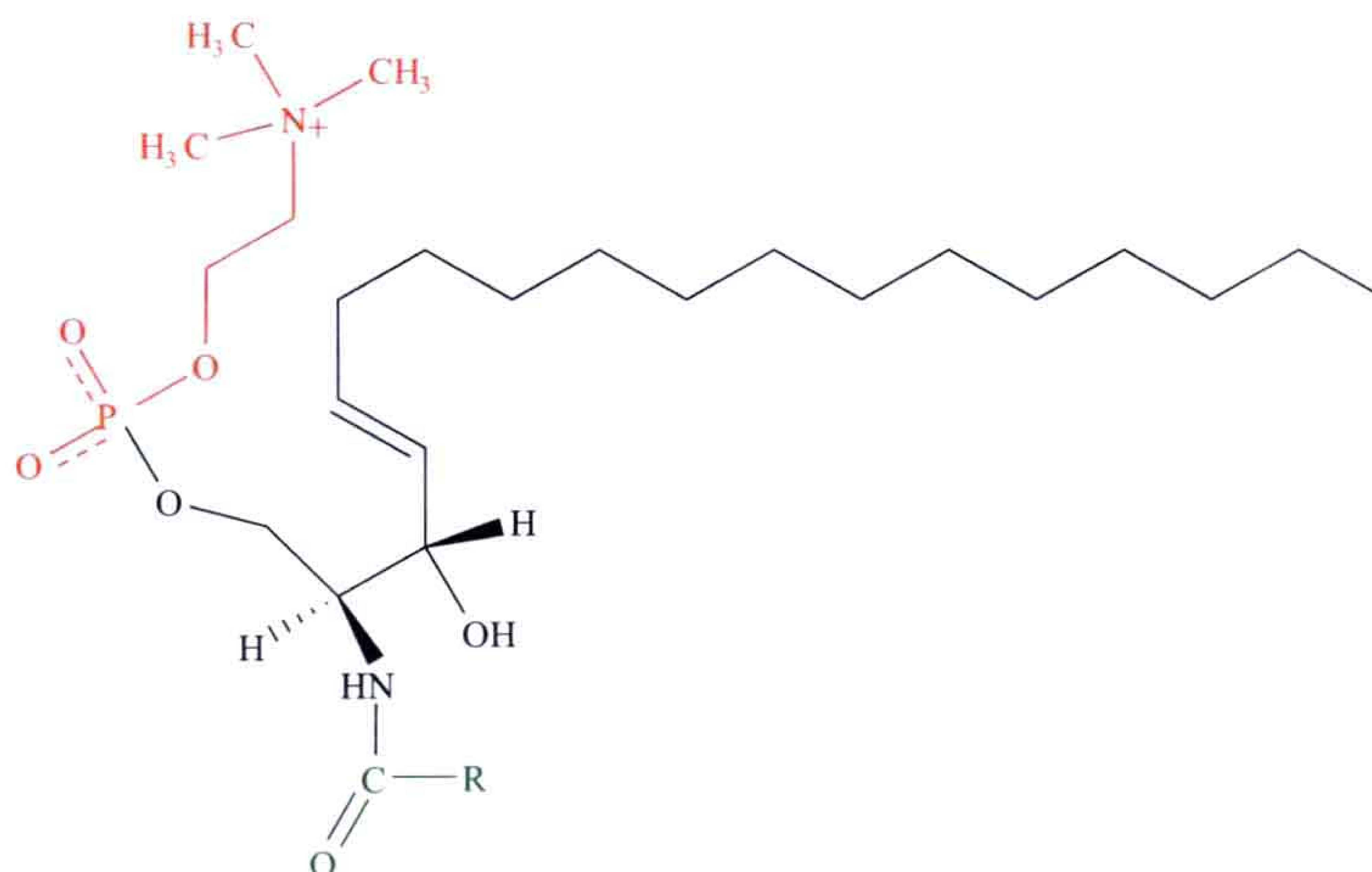


图 4.6　鞘磷脂的组成成分示意图。红色部分是磷脂酰胆碱；黑色部分是神经鞘氨醇；绿色部分是脂肪酸；R 代表脂肪链。没有磷脂酰胆碱基团的这种脂质被称为神经酰胺。

磷脂中含有多种不同的极性头部基团。例如，胆碱和乙醇胺使磷脂的头部基团在中性 pH 时为两性离子，而丝氨酸、甘油和磷酸基团使头部基团带负电荷。

磷脂酰胆碱及相关的磷脂通常在 sn-1 位含有一个饱和脂肪酸，而在 sn-2 位置通常是一个包含 1~6 个双键的不饱和脂肪酸。磷脂 sn-2 位置酯键的水解会生成 1- 酰基 -3- 磷酸甘油，又被称为溶血磷脂。它可以像强力表面活性剂或者去垢剂一样导致细胞裂解。例如，一些蛇毒中就含有可以生成溶血磷脂的磷脂酶。

另一类磷脂含有六羟基环已烷，其又称肌醇。磷脂酰肌醇在所有真核生物的膜中都存在，对于调节细胞对激素及其他外部因子的反应有特异性的作用。同时它还可以作为使特定蛋白定位到膜表面的锚结构的一部分（附录 E）。

细菌和植物常合成阴离子型磷脂酰甘油，其极性头部的磷酸基团将第二个甘油酯化。细菌与线粒体一样含有双磷脂酰甘油，其又称性磷脂，它在甘油的 1 位和 3 位各含有一个磷脂酰基基团。

在嗜盐的、嗜热的及产甲烷的细菌中存在的脂质大多是磷脂或者糖脂。支原体如莱氏无胆甾原体（*Acholeplasma laidlawii*）中最主要的膜脂是糖酯、单葡萄糖二酰基甘油（MGlcDAG）和双葡萄糖二酰基甘油（DGlcDAG）。

纤毛虫原生动物如四膜虫及一些其他的无脊椎动物中有大量的含有碳－磷（C-P）键的磷脂。在这些脂质中，磷酸乙胺取代了磷酸乙醇胺，使其可以抵抗磷脂酶 C 的水解。

四膜虫外膜中的磷脂在其 sn-1 位也变成了带有烷氧基的醚脂类，这让它们也具有磷脂酶 A_1 的抗性。这两种特点似乎可以保护原生动物裸露的细胞膜不受它们自身分泌到环境中的磷脂酶的影响。另外，在某些海洋藻类中还发现一种特殊的含砷磷脂，被称为氧－磷脂酰三甲基砷乳酸。

4.2.2 液晶相的膜脂

脂质和蛋白质是细胞膜的主要成分，它们之间紧密的相互作用越来越被认为是生物膜行使功能的基础。长久以来人们一直认为脂质仅仅起到“惰性基质”或者“油质膜层”的配角作用，只是作为蛋白质的一种二维溶剂而存在；而现在，人们认为脂质的作用比这个重要得多。脂 / 水系统具有非常丰富的多形性，迄今已经发现了 10~15 种不同的相结构。多种学科对脂质可以形成液晶相的现象表现了浓厚的兴趣，主要包括胶体化学、生物化学、材料科学及结构生物学。与多种不同“组”学相应的、一种新型的、被称为脂质组学的科学已经出现。

胶体或者脂质系统相图的知识是非常重要的。一般来说，绘制一幅完整的相图是一件冗长而且乏味的工作。基本的相图绘制方法是简单地将成分混合到一起，然后宏观上观察形成的相的数量及本质。最为直观也是最为通用的方法是直接肉眼观察或者使用偏振显微镜观察。只有当不同的相位可以在宏观上分辨以后，才能容易地决定相图。然而，脂 / 水系统通常非常黏稠，因而需要相当长的时间才能达到不同相位分开在不同区域的状态。对于这些系统来说，平衡的相位经常仅仅会以几十到 100μm 的大小分散交错存在。对这种系统来说，视觉的观察是远远不够的，而需要使用其他的方法。三种最为常见的方法分别是量热法、X 射线散射及光谱的方法（包括核磁共振和荧光等）。其中，量热法仅仅可以应用于绘制温度 - 成分（T-X）的相图。差值扫描量热法可以测量热容量对应温度的函数。X 射线小角散射（SAXS）可以用于测定组成相位的聚集体的总体结构。光谱方法测量的通常是对局部分子环境敏感的参数。因此，同样的分子位于不同的相位，在使用光谱法测量时可以给出不同的信号。在一个多相位的样品中这种特点非常有用，利用它可以计量相位数量，并且通常可以揭示相位所处的物理化学环境的一些特点。其中，^{2}H 和 ^{31}P 的核磁共振是非常有效的方法。

一些有关脂质的有趣技术应用包括利用脂质材料自组装的特性使其形成不同的结构，并且利用这些脂质微结构作为制造稳定的沸石状复合材料的模板。最近还有利用液晶相脂质作为生物容性介质的技术，可以用于药物传递。脂质学的这些进步引起了人们巨大的兴趣，许多商业企业都对这些在医药和化妆品领域有着多种应用价值的材料进行大量的投资。

吉布斯相律（Gibbs phase rule）及相位特性的描述

在水中散布的脂质既呈现出温度的多态性，又呈现溶解度的多态性。用于描述相

位平衡的最重要的热力学原理是吉布斯相律，其公式是

$$F+p=c+2 \tag{1}$$

式中，F 表示当我们将所有约束条件考虑在内后剩下的自变量数目的自由度（如温度、压力和摩尔比等一些与物质的量无关的变量）；p 表示平衡状态下共存的相位个数；c 是成分数目。在恒压条件下，$F= c-p+1$。在一个脂 / 水的二元系统 [如图 4.7 所示的二棕榈酰磷脂酰胆碱 (DPPC)/ 水组成的系统] 中，$c=2$，而 $F=3-p$。因此，两个热力学的变量，如温度和摩尔比在只有一个相位的区域中可以自由变换（图 4.7）。

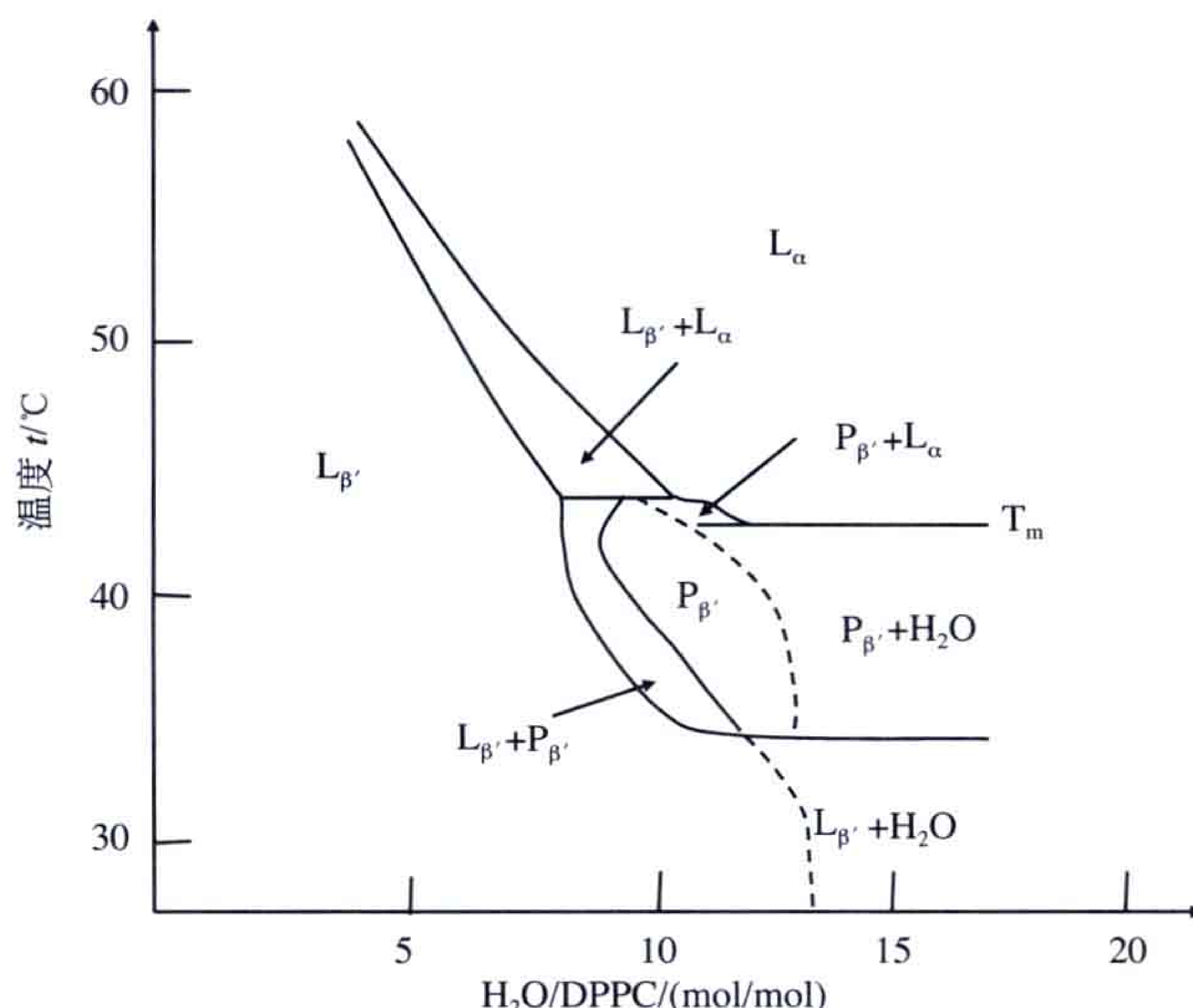


图 4.7　DPPC 和水的部分相图。在低温时会形成凝胶状的 $L_{\beta'}$ 相位，在高温及较高的水含量时则会形成稳定的薄层状液晶 L_α 相位。在相图的中间 $P_{\beta'}$ 相位仅在狭窄的温度和水相中稳定 [经授权根据参考文献 Ulmius J, Wennerström H, Lindblom G, Arvidson G. (1977) Deuteron NMR studies of phase equilibria in a lecithin-water system. *Biochemistry* 16: 5742-5745. Copyright (1997) American Chemical Society 改编]。

通常，成膜脂质在整个浓度的变化范围内都会形成一种薄层状（L_α）相位（表 4.2），此时 F=2。因此，温度和脂 / 水组成在这种单一相位的区域中可以发生变化。然而，在过量的水中，L_α 相位可以与“纯”水形成平衡，也就是 p=2 时，F=1，即只有温度可以发生变化。在含有过量水的系统由胶体相位（P_β）向 L_α 相位转变时，p=3、F=0，此时，系统是不变的。因此胶体相位和 L_α 相位只能在特定的温度下共存，这个温度被称为主相变温度 T_m。注意，含有不同脂肪链长度及不饱和性的膜脂质混合而成的物质，由于有着大量不同的组分，并不能形成一个真正的二元系统。然而，总体来说如果我们把所有的具有细微差别的膜脂质都看成一种，并不会对我们绘制相图产生任何的困难，但是始终要记住的是在这样一个“假二元系统”中，可能会出现“出乎意料”的结果。

对于恒压条件下的三种组分，F=4−p，如果我们想绘制一个二维的相图，就需要将温度固定。因此，对于三组分系统，我们使用一种三角形的图表来绘制相图（参考其他物理化学教科书），三角形的三个顶点分别表示一种纯的组成分。平衡相位的最大数目是三、四元相图的一个典型特征是其三相分界的三角行区域，这与二元系统的三相分界线相类似。

在绘制相图的过程中，所谓的“层级定律”是非常有用的。在相图两相区域中的一个点，不只定性的表示两种相位的存在，而且可以定量的表示每一种相位的相对含量。平衡的两种相位的相对含量是由这一点在两平衡相成分点间连线与其各自的相边界线的相对距离决定的，这就是层级定律。对于二元系统来说，两平衡相成分点间的连接线总是水平的；而对于一个四元系统，这种连接线的方向的预测则并不简单，只能用实验的方法确定。此时，NMR 的方法最为方便，因为 NMR 波谱中的一个峰值下的面积是与给出信号的原子核的个数或者分数成正比的。这可以用于确定所研究样品中不同相位的比例。

以下的命名规则在相图中会经常遇到。大写的拉丁字母描述长程有序的类型（一、二或三维格子），下标的系列字母表示的是有序的（β）或者无序的（α）脂肪酰链。下标的罗马数字 I 和 II 分别表示普通的及反转的液晶状结构。

表 4.2 和图 4.8 列出了最为常见的膜脂质液晶状相位。

表 4.2　膜脂质形成的液晶相位和凝胶相位

相位	维度	命名	规则性
薄层	一	L_α	不规则，液相
波纹胶状	二，斜的或者居中的	P_β	波纹状
凝胶	一	L_β	全反式 * 酰基链
普通六角	二	H_I	不规则，液相油在水中
反向六角	二	H_{II}	不规则，液相水在油中
立方	三	I	不规则，液相
普通立方	三	I_I	不规则，液相
反向立方	三	I_{II}	不规则，液相

* 这里全反式表示酰基链的构象（见任何有机化学教科书）。

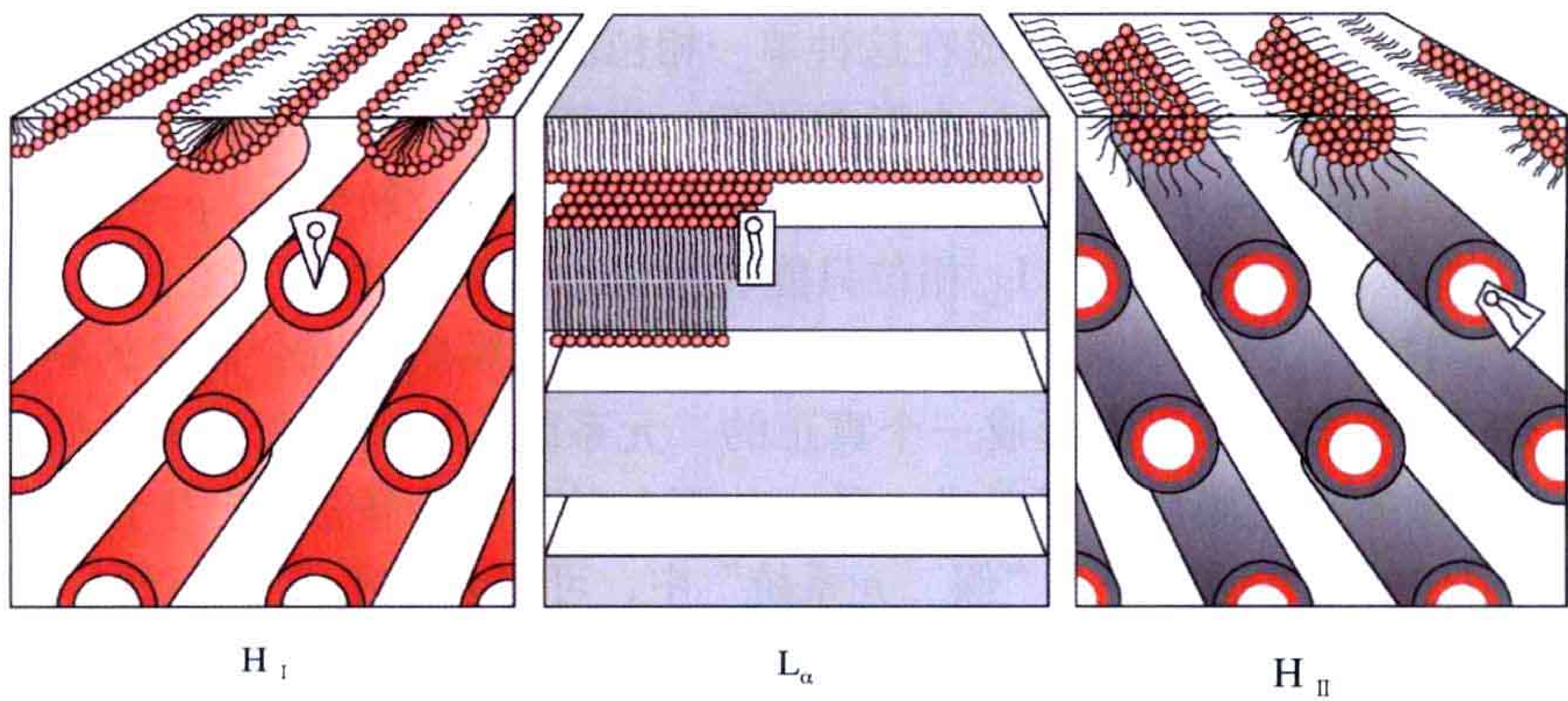


图 4.8　一些常见的液晶相位结构。从左到右（即按照水含量降序排列）：普通六角相位 H_I、薄层相位 L_α、反向六角相位 H_{II} [经授权根据参考文献 Lindblom G, Rilfors L. (1989). Cubic phases and isotropic structures formed by membrane lipids—possible biological relevance. Biochim Biophys Acta 988: 221-256. Copyright (1989) Elsevier. 改编]。

如图 4.7 和图 4.9 所示，降低水含量可以引起相位由液晶相向胶体相或者在不同的液晶状相位之间的转变。这是由于磷脂脂双层通过短程的排斥力相互作用，水含量越低，这种相互作用就越强。这种聚集体间（脂双层之间）的相互作用可以导致相变。因此，脂双层之间的作用力与相位行为之间有着紧密的关系。降低水含量时，脂双层倾向于转变成排斥力较弱的情况。在相变时，脂双层之间的相互作用可以补偿脂双层之内的相互作用的差别。

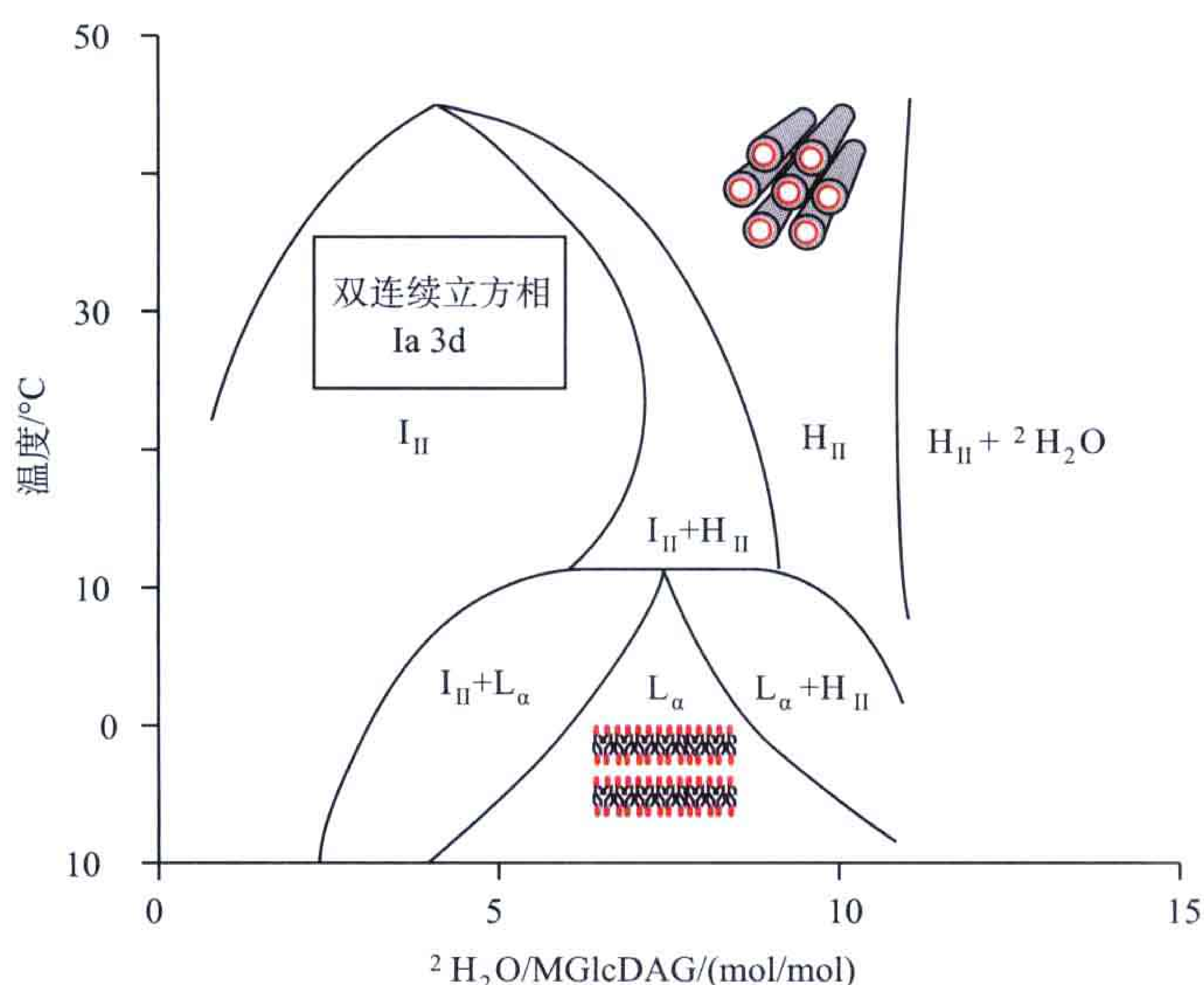


图 4.9　单葡糖糖酰双甘油（MGlcDAG)/ 重水系统的部分相图。其中 MGlcDAG 来自细菌莱氏无胆甾原体。注意 L_α 相位仅在远远低于生长温度的低温下稳定，也就是说在生长温度下只能形成 H_{II} 和 H_{III} 这两种非薄层相位 [经授权根据参考文献 Lindblom G, Rilfors L. (1989) Cubic phases and isotropic structures formed by membrane lipids—possible biological relevance. Biochim Biophys Acta 988:221-256. 改编]。

4.3　两亲分子自组装成不同的聚集体结构

4.3.1　脂质组装和自发弯曲

一种典型的脂质结构如图 4.10 所示。由于极性的脂质分子是由亲水和疏水两部分经由骨干残基连接而成，所以又被称为两亲分子。在水中，两亲分子可以组装成不同的聚集体并形成相位。为了定性理解两亲系统中相行为，包装参数是一个最为重要的概念，它的定义基于脂质分子的几何构造或者整体形状（图 4.11）。脂质分子的自组装取决于这个无量纲的包装参数，其定义是

$$P=v/al$$

式中，v 是易变的烃基链的体积；l 是疏水链的长度；a 是头部基团的最大横截面面积，如图 4.11 所示。

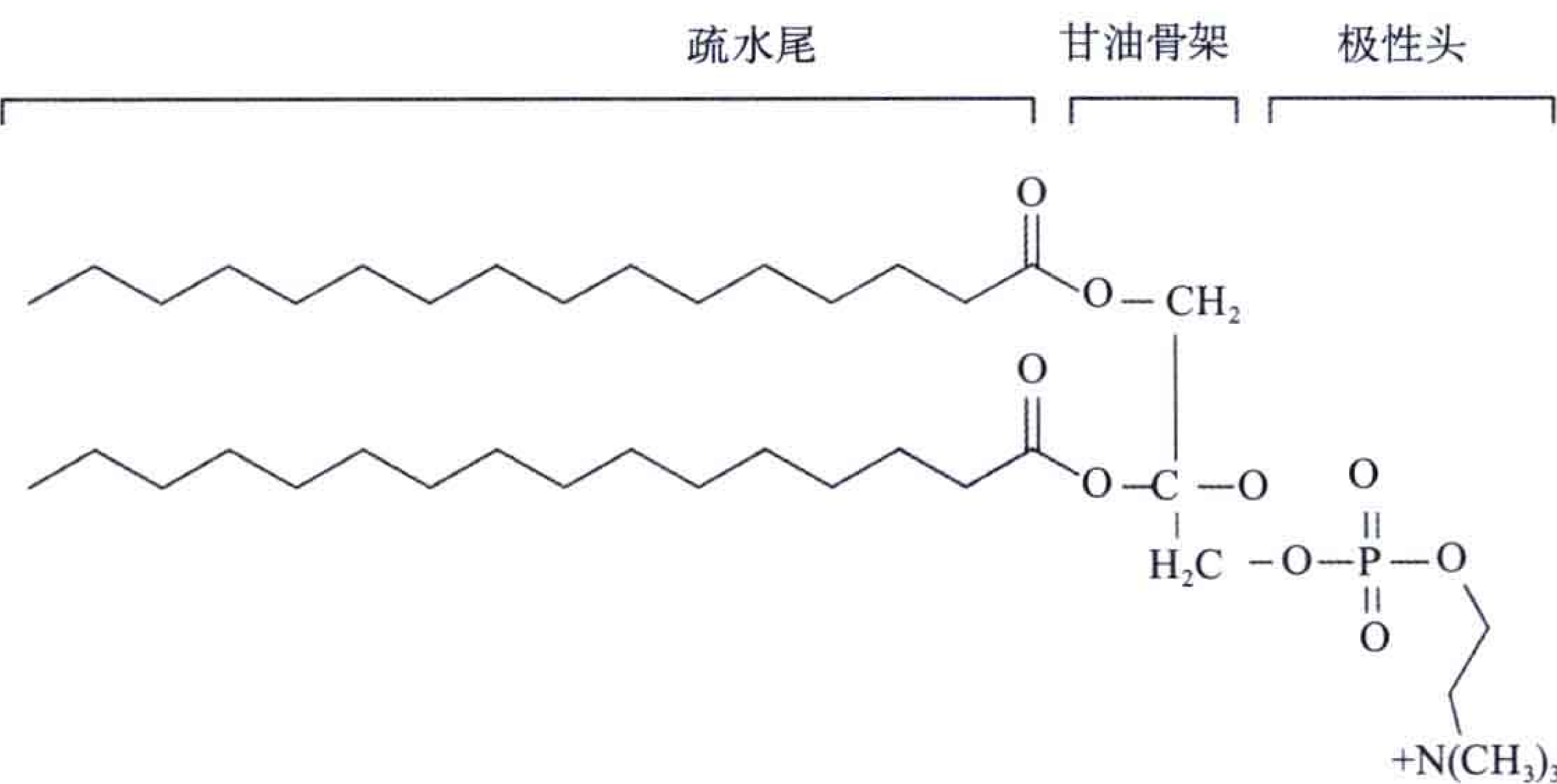


图 4.10　一种典型的脂质、磷脂（磷脂酰胆碱，PC，俗称卵磷脂）。它的两亲特性表现在疏水的烃基酰链（尾部）和与骨架（此情况下为甘油）相连的亲水极性头部基团。

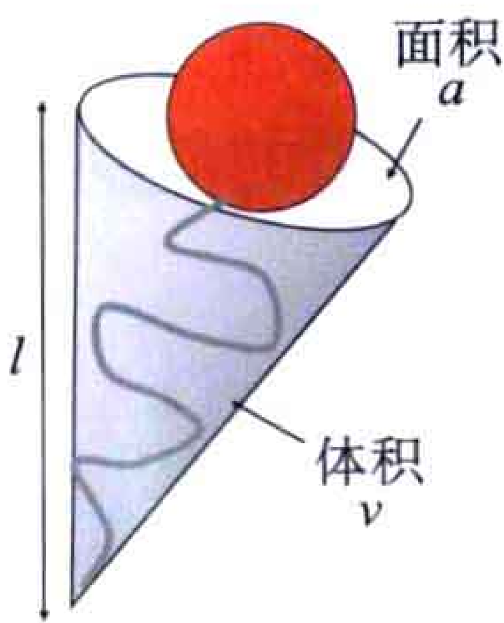


图 4.11　图 4.12 中形成球状微团的脂质分子的示意图。红色小球是极性头部基团，疏水性的尾部用灰色表示。用来定义包装参数的成分，如图所示。

当包装参数（有时也被称为表面活性剂数）等于 1 时（圆柱形的分子，图 4.12），有利于分子形成脂双层的结构。而如果 $P>1$，脂质分子呈楔形的结构，单层的脂质倾

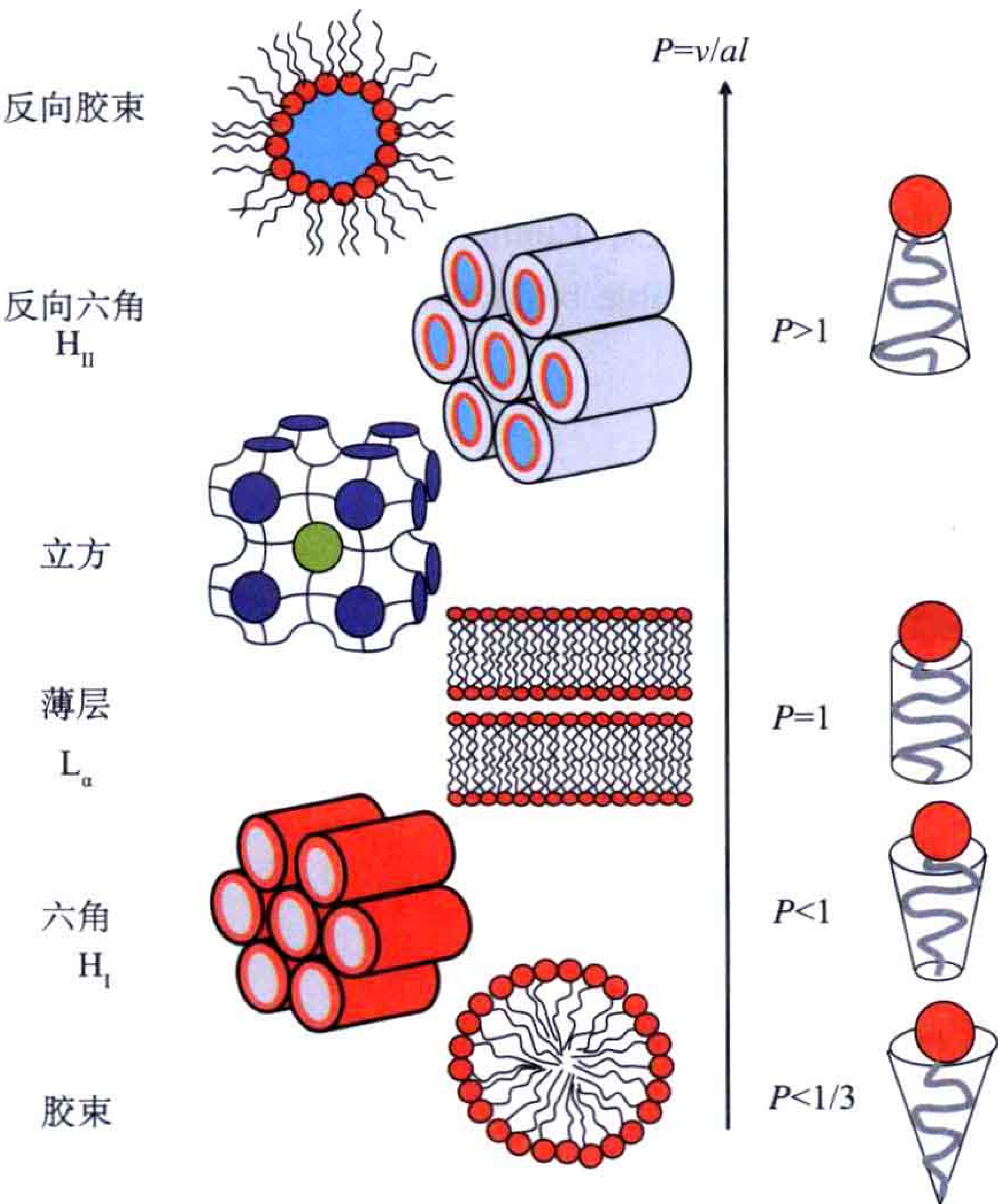


图 4.12　具有不同形状和包装参数的脂质分子。在不同相位（微团溶液以及液晶相位）时可能形成的聚集结构如图左侧所示。液晶相位具有液体的特点，同时像晶体一样具有长程的规则性。

向于向水的方向弯曲，即形成反向的微团或者 H_{II} 型液晶状相位（图 4.12）。

这种简单的方法在解释某些特定分子形成的聚集体形状方面非常有用。然而，需要提醒的是，表面积 a 在不同温度、电荷、浓度的情况下的计算方法非常复杂，需要细致的考虑。例如，在较高水含量时，当向表面上是圆柱形的脂质分子卵磷脂（phosphatidylcholine, PC）所形成的似乎是稳定的薄层状液晶状相位中添加烷烃或者疏水性多肽时会形成反向的六角形 H_{II} 型相位，而分子形状的理论并不能充分解释这种现象。显然，这应该说明即使多层结构中卵磷脂分子的包装系数稍大于 1，也会有其他的因素限制其形成弯曲的单层状结构。究其原因则是，将 PC 分子组装成大的 H_{II} 型相位的圆柱体首先需要在其间隙中生成大的真空，下面的讨论将会说明这是不可能的。

这种 PC 分子形成的脂双层是“无效的”（见后文）。这可以用一个被称为脂单层曲率的概念来解释，脂单层曲率的概念与包装参数相关，但是有更为广泛的应用范围，而不仅仅只与脂质分子有关。将膜变形所需要的能量是由膜的结构和弹性所决定的。没有变形的、无应力的膜状态被称为自发态。膜的弹性性质取决于自发态直接的偏差、这些偏差需要的力及新形状中积累的能量。

为了理解这些，我们首先简单回顾一下弯曲的膜物理化学性质和牵扯的热力学知识。首先，我们需要做出一些定义。在三维空间内的一个薄层上的任意一点，都可以定义两个主要的曲率半径 R_1 和 R_2，以及局部的曲率 $c_1=1/R_1, c_2=1/R_2$（图 4.13）。

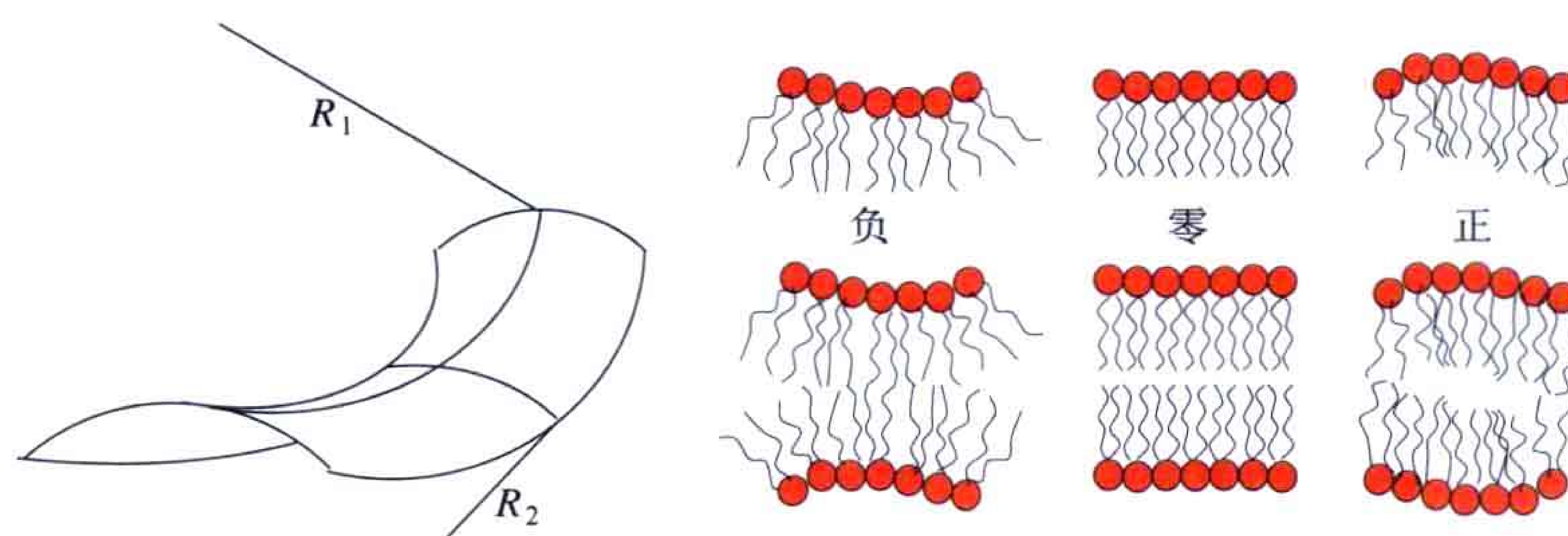


图 4.13　左：两个膜曲度半径的定义，在本图中马鞍状的表面其两个半径符号相反，右：定义曲度半径正负符号（按照惯例）的图解。

曲率的正负是可以随意定义的，按照惯例，人们通常使用图 4.13 中的定义方式，从双层膜的内部包裹的区域向周围介质的方向“向外”凸起的部分具有正的曲率。因此，球状的微团具有均匀的正曲率，因为其曲率半径 R_1 和 R_2 相等且都是正的。在双连续性的立方相位结构（图 4.15）或者在出芽小泡颈部发现的马鞍状膜的曲率在一个主轴方向是正的而在另一个方向则是负的。将单层结构弯折每单位面积所消耗的能量可以使用吉布斯弹性曲率能量（Gibbs elastic curvature energy）描述，它是两项的和，一项由单层结构的总曲率（c_1+c_2）所决定，另一项则是由 c_1 和 c_2 的乘积决定：

$$g_c=1/2k\{1/2(c_1+c_2)-c_o\}^2+k_G c_1 c_2 \tag{2}$$

式中，k_c 是弹性弯曲常数；k_G 是鞍状曲面（又称高斯曲面）常数；$c_o=1/R_o$ 是自发曲率，即松弛无张力脂质单层结构的半径 $R_{o'}$。自发曲率是一种测量脂质单层弯曲形成非平面几何结构的趋势的量（总弹性曲率能量 G_c，是将 g_c 在整个单层结构上进行积分的结果）。

由公式可见，使平均曲率 $1/2(c_1+c_2)$ 尽量接近自发曲率 c_o 从而使曲率的自由能最小化是有利的。在 H_{II} 型相位结构中，脂质单层弯折形成半径为 R 的圆柱状结构。研究发现，形成 H_{II} 型相位所需的吉布斯自由能包含两个部分。第一部分是弯曲部分的能量，由于在 H_{II} 型相位中，脂质分子的酰基链必须伸展才能填满邻近的圆柱体之间形成的疏水性区域（图 4.14 中的绿色区域），因此第二部分能量是非零的组装能量 g_p。总的吉布斯自由能就是弹性曲率能量和组装能量的和，$g_{tot}=g_c+g_p$。因此我们得到两种相反的物理作用力，即曲率作用和组装作用共同作用的情况，这种情况就是前面提到的所谓“无效的”情况。研究发现，这种“无效的”情况在加入疏水性分子如烷烃时会减少或者消除。在 H_{II} 型相位中这些加入的分子会优先分配到圆柱体之间的烷烃区域（绿色区域），以此来填充在生成 H_{II} 型相位时产生的空隙体积（图 4.14）。

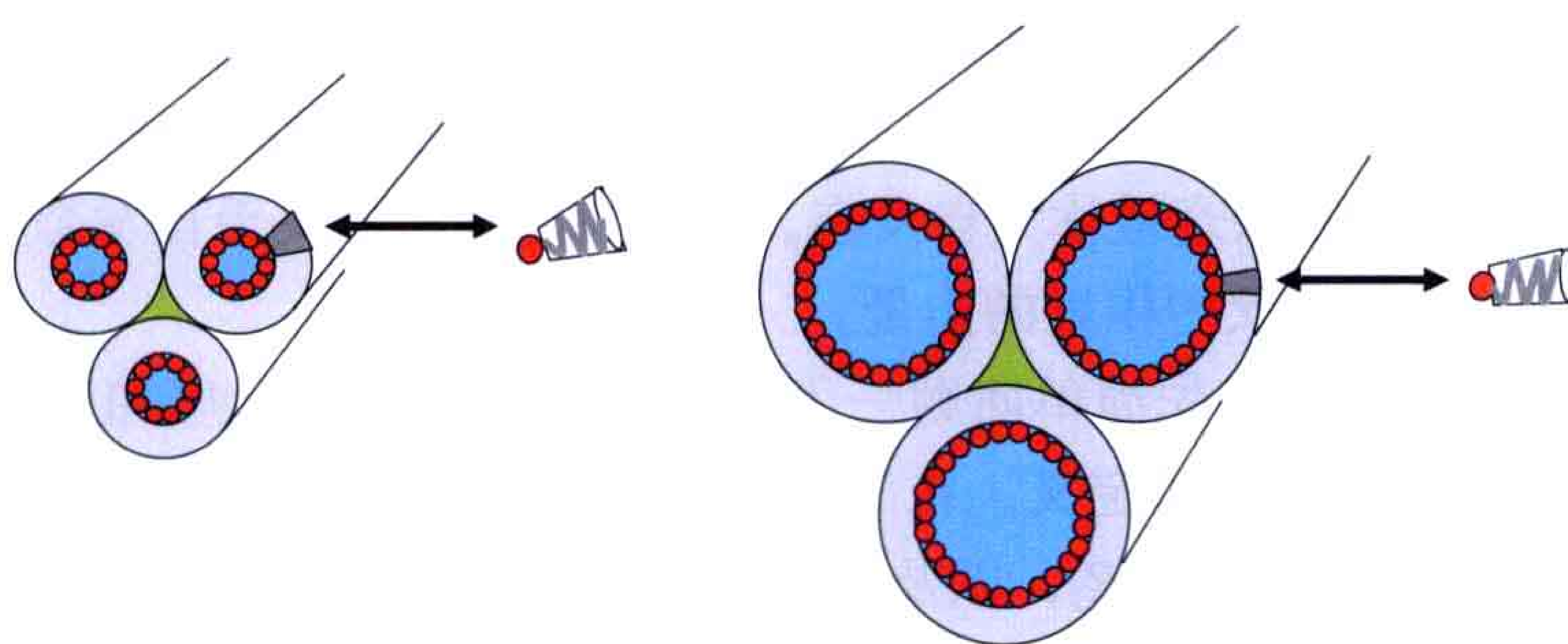


图 4.14　当脂质分子（放大的）形状变得更加不像楔状，组成 H_{II} 相位的圆柱体曲率半径升高，H_{II} 圆柱体中吸收的水量随之升高。绿色的中间孔隙的体积也会升高，此时 H_{II} 相位仅会在圆柱体之间含有疏水分子如烷烃时形成。注意有饱和脂肪链的脂质比有非饱和脂肪链的脂质在 H_{II} 相位时会形成较大的圆柱体 [经授权根据参考文献 Sjölund M, Rilfors L, and Lindblom G. (1989) Reversed hexagonal phase formation in lecithin-alkane-water systems with different acyl chain unsaturation and alkane length. Biochemistry **28:** 1323-1329. Copyright (1997) ACS. 改编]。

在双连续性的立方相位中还发现了另一种由弯曲的单层膜组成的、迷人的液晶状结构。这里单层膜在每一侧面都保持最小表面（在最小表面上的每一点的平均曲率接近于零，即 $c_1+c_2=0$）。这一表面处于双层膜的中间部，而不是在其与极性或疏水区域的交界处。在这种结构中，脂单层由于立方相的这种结构，与 L_α 或者 H_{II} 相位相比较，立方相基本不存在“无效的”情况，因此立方相经常出现在脂质相图 L_α 和 H_{II} 相位之间的位置。三维的立方相结构被看做是膜蛋白结晶的一种可能基质。

通常，在考虑双连续立方相位的稳定性时，曲率能量起到重要的作用。另一种主要在很多非离子型双亲分子 - 水系统中发现的液晶状相位是各向同性的 L_3 相位（经常又被称为海绵相位）。海绵相位一般仅在很窄的温度和组成范围内比较稳定。与许多立方相位一样，海绵相位的基本结构单位也是由相互连接的脂双层形成的网状结构。使脂质形成 L_3 相位而不是 L_α 相位的驱动力是 L_3 相位可以让脂单层有形成最优化曲率的机会。L_3 相位的结构来自熔化的或者混乱的立方结构（图 4.16）。在脂双层之间存在较弱的相互作用时有利于形成这种混乱的结构。在高的脂质浓度及强的脂质双层间作用力下，立方相位可以与 L_3 相位形成一种平衡态。

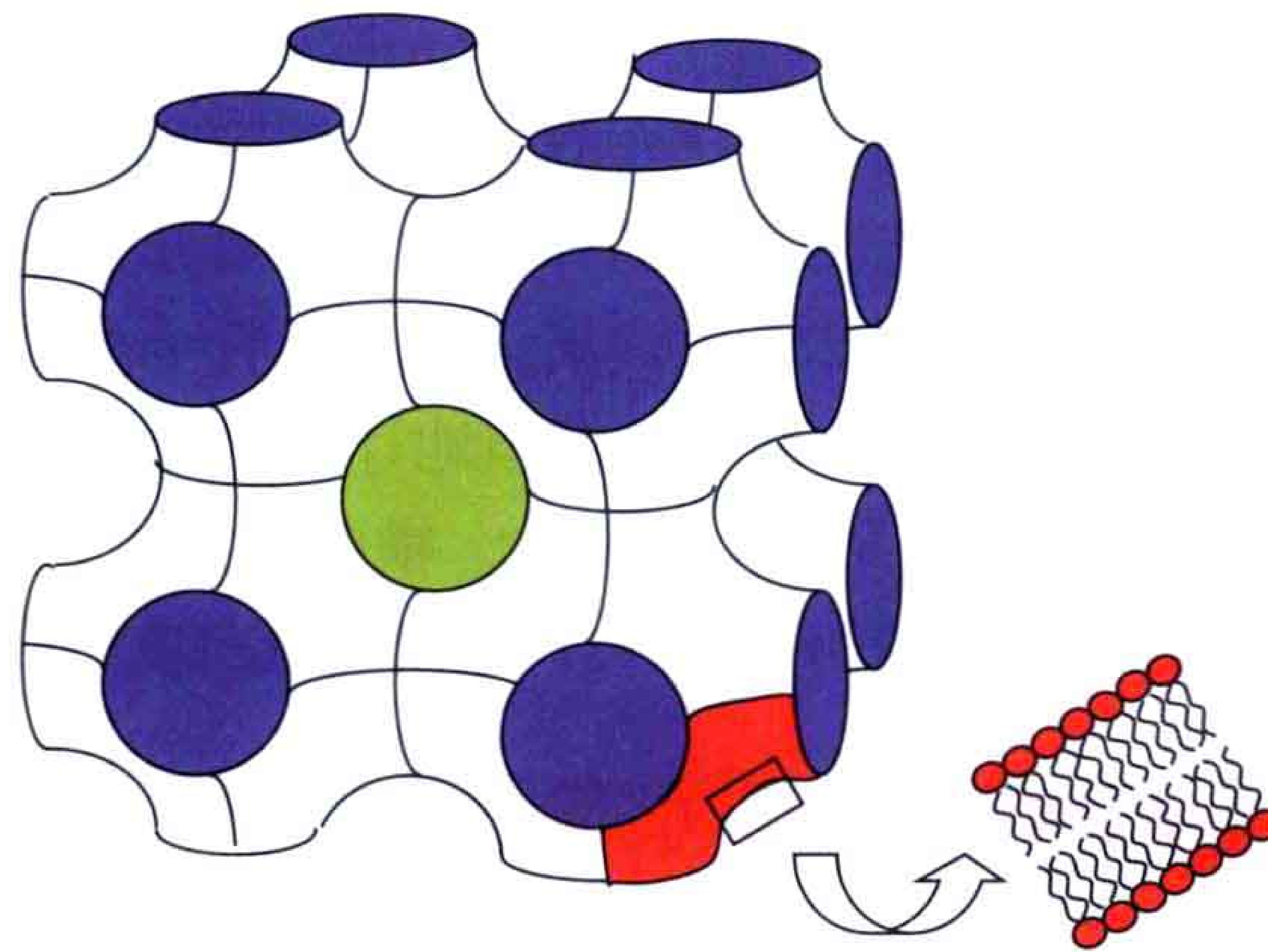


图 4.15　一种双连续立方相位（有时也被称为水管工的梦魇）的结构示意图。注意脂质单层依附在最小表面（Schwarz 表面）的内外两面形成“皱褶的”脂质双层。这个相位中有两个独立的“水系统”，在蓝色区域的水分子永远无法穿过脂双层到达绿色的区域 [经授权根据参考文献 Lindblom G, Rilfors L. (1989) Cubic phases and isotropic structures formed by membrane lipids—possible biological relevance. Biochim Biophys Acta **988:** 221-256. Copyright (1989) Elsevier. 改编]。

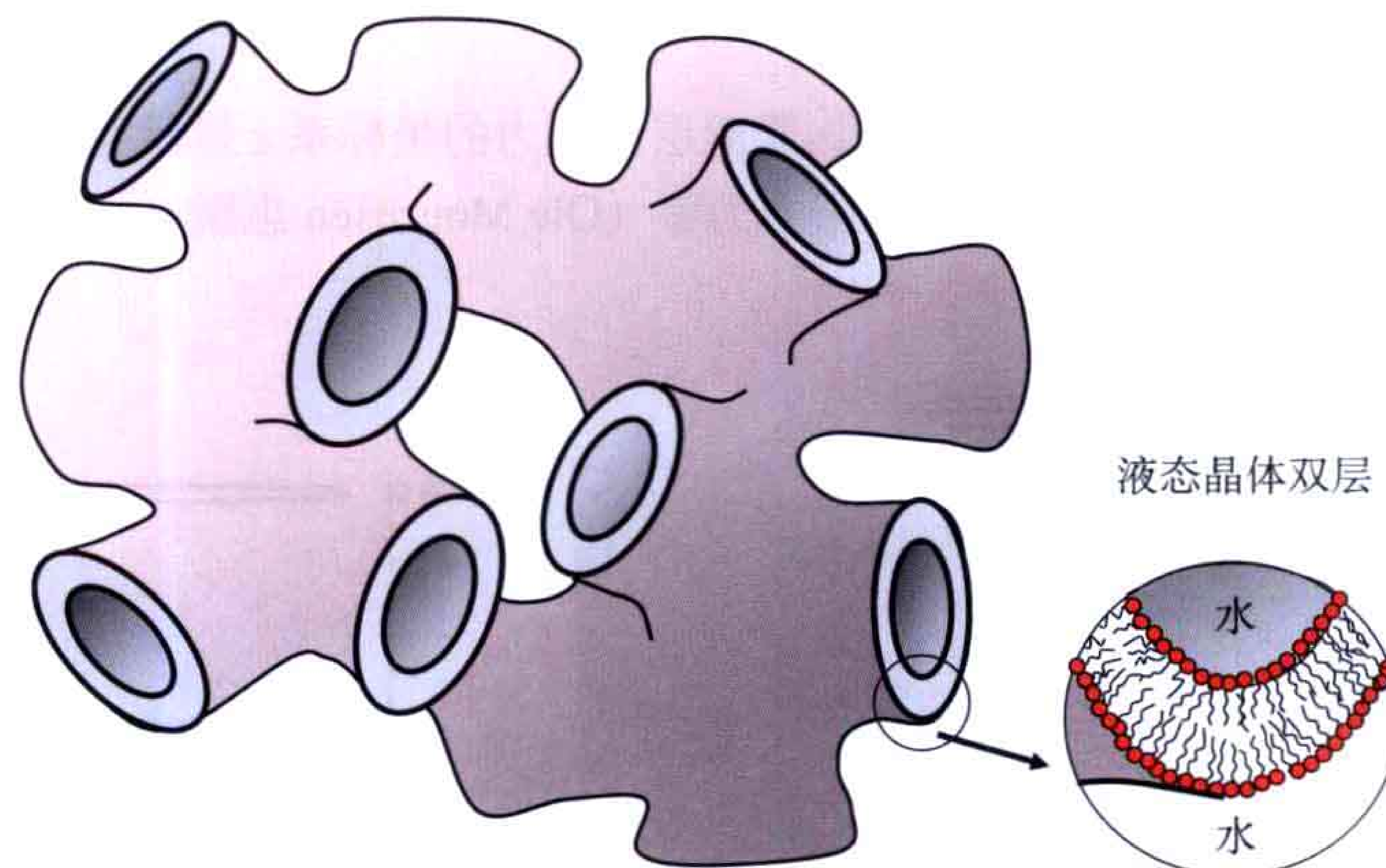


图 4.16　海绵相位（L_3 相位）的示意图。

4.3.1.1　脂质组装和横向压力

脂双层另一个重要的物理化学特点就是其横向压力。如前所述，使脂质脂肪链远离水分子的疏水作用最终使得脂质分子聚集形成脂双层。然而如前所述，由于被限制在会出现无效结构的脂双层结构中，脂质分子经常受到巨大的应力。这带来了脂双层的一个最重要的物理特征，即其横向压力外形（图 4.17）。

从图 4.17 中可以推断，在稳定脂双层聚集体的过程中牵扯到不同的力。当然，对于形成平衡状态的脂双层，各种作用力相互抵消使其净合力为零。由于各种力作用于不同的平面，在双层膜的各个部分的压力分布也是不均匀的。横向压力外形包含三方面的贡献，即头部基团之间的排斥作用形成的正向压力、疏水 / 亲水表面的界面张力所形成的负向压力，以及柔性的脂质烃链熵排斥造成的正向压力（参照 4.3.1 节关于向脂

双层中添加烷烃使相位向 H_{II} 相转变的讨论）。因此，构成脂双层的脂质决定了其横向压力外形，即具有大的脂肪酸侧链的脂质会比含有饱和直链脂肪酸的脂质造成更大的链压力。由于脂双层非常薄，两个界面之间的大的界面张力需要分布在很短的距离之内。这意味着脂质链带来的反向作用的压强非常大，典型的是几百个大气压。脂双层每个表面的界面张力（γ）约为 $50 mN \cdot m^{-1}$。由此可以简单地计算出脂双层内部的横向压力需要抵消此在脂双层厚度（d=2.5~3 nm）距离上的张力。横向压力等于 $2\gamma/d \approx 40\ 000$kPa，约 400 个大气压。这种巨大的压力可以影响膜蛋白，如改变蛋白质结构等（图 4.18）。

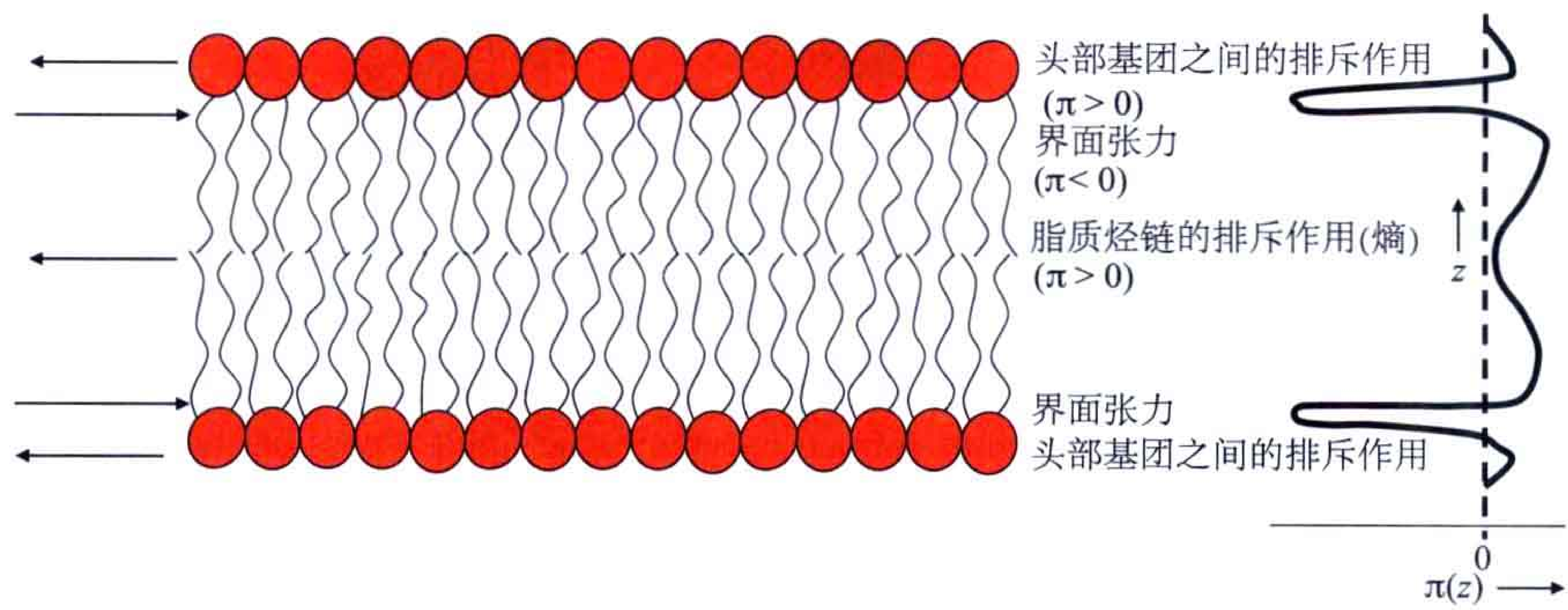


图 4.17　横向压力 $p(z)$ 在脂双层上的分布图。脂双层中压力的坐标系 z 显示在图的右侧。脂双层中间的横向压力可能非常高，但是脂双层上总的压力为零（Ole Mouritsen 惠赠）。

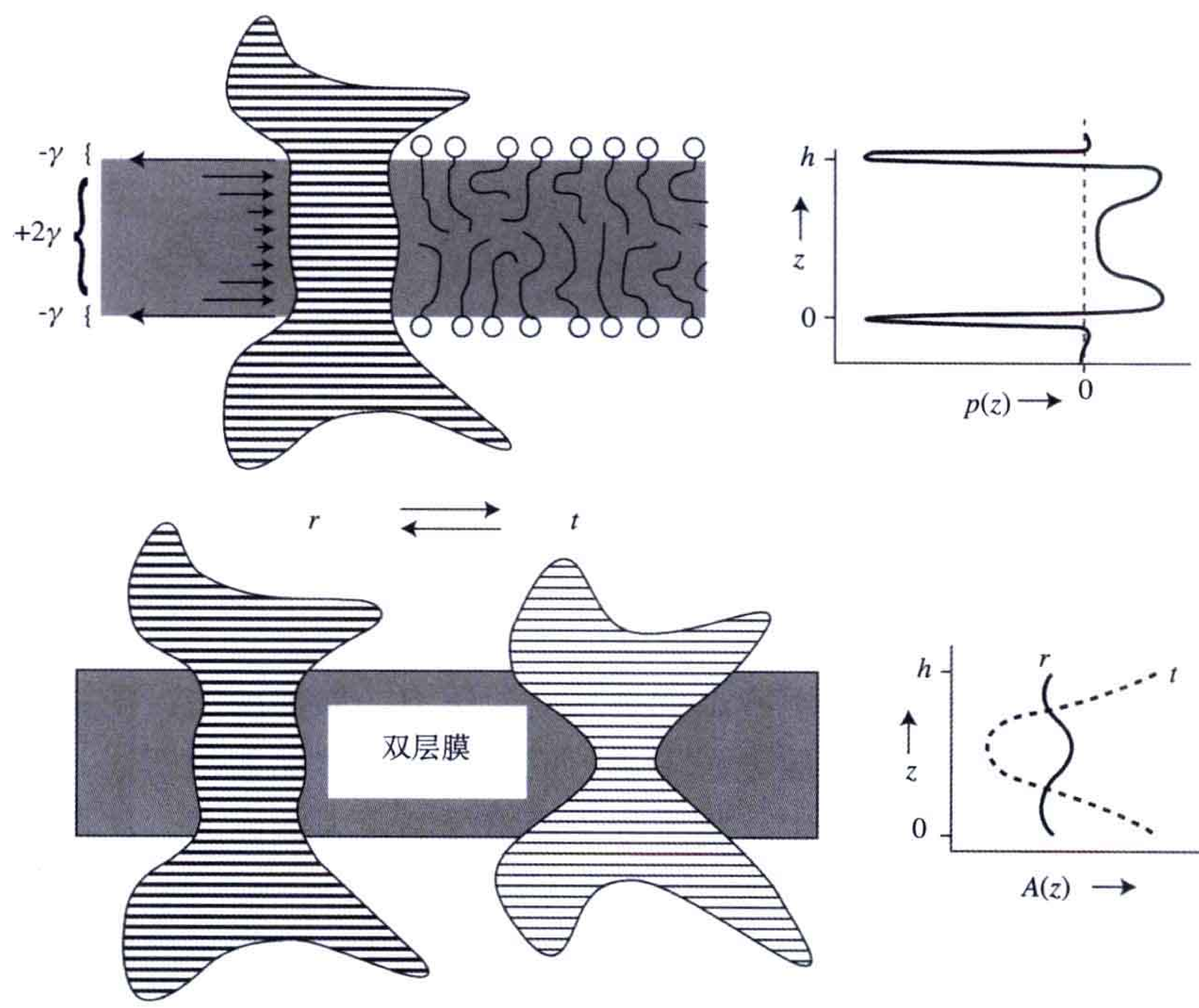


图 4.18　横截面 $A(z)$ 显示高的横向压力 $p(z)$ 可以使内在膜蛋白（条纹或者虚线）的构象发生变化。蛋白质可能处于 r 和 t 两种状态中的任意一种，γ 是界面张力 [经授权根据参考文献 Cantor RS. (1997) Lateral pressures in cell membranes: a mechanism for modulation of protein function. J Phys Chem B 101: 1723-1725. (Copyright (1997) ACS.) 重印]。

与上述内容类似，向双层膜中引入不同形状的脂质可以改变横向压力，使细胞膜上的通道或者孔道处于开或者合的状态。这样的通道被称为机械敏感通道。机械敏感通道可以作为膜包埋的机械电力开关，可以感应脂双层的变形而开放大的充满水的孔道。这个过程对于生物感受直接的物理刺激信号，如触觉、听觉和渗透调节是极为重要的。例如，如图 4.19 所示的大的原核生物压力敏感通道（MscL），它在开放的状态下呈现高度变化的状态，可以支持大至 25Å 的充水孔道。当溶血磷脂酰胆碱（LPC）被加入细胞膜时，这个通道会打开，显著地降低其激活阈值，并减少蛋白质的横向压力（图 4.20）。

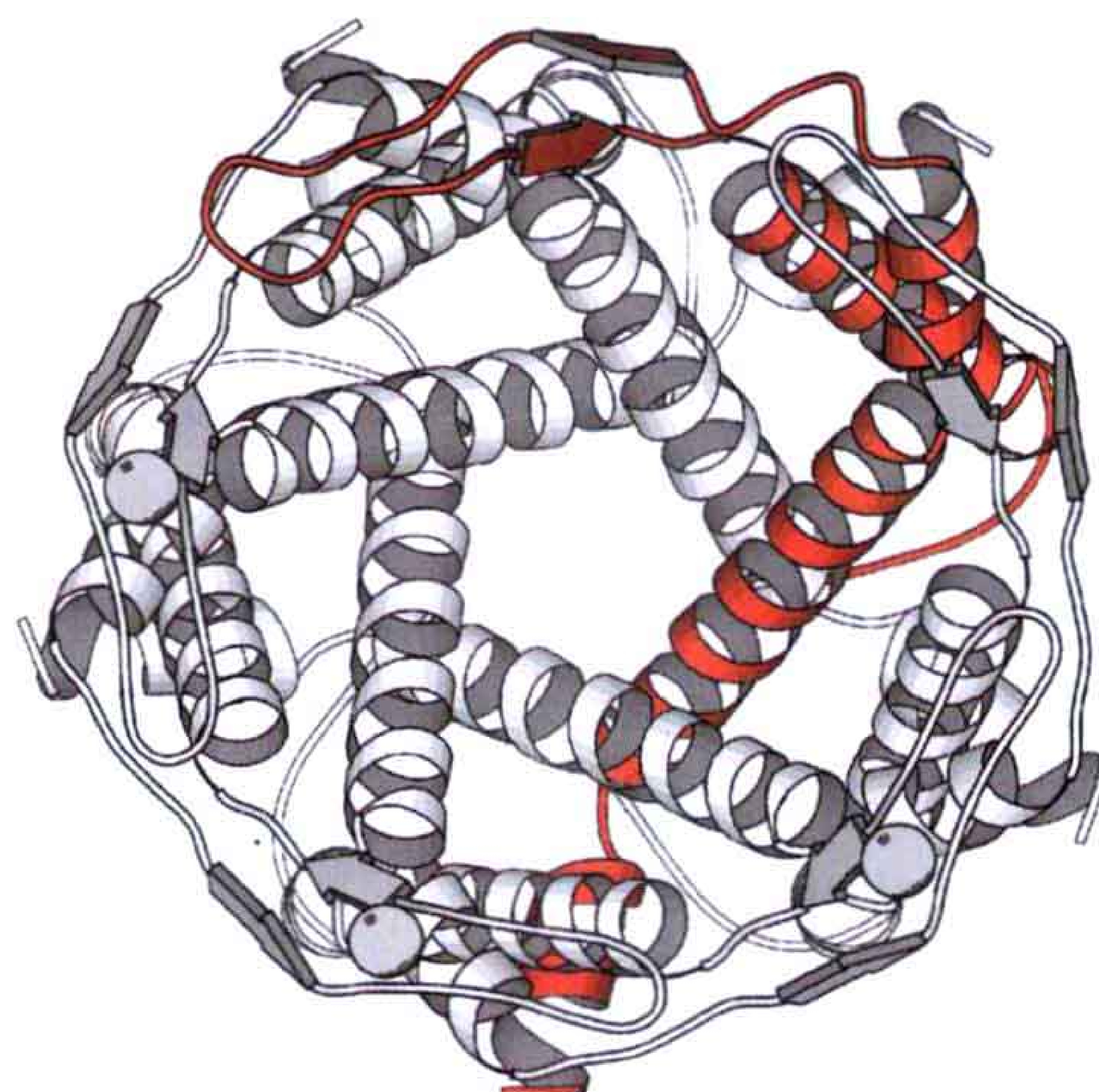


图 4.19 位于结核分支杆菌外膜的大型膜通道 MscL 闭合状态的结构。图为与膜表面垂直的视角。5 个亚基构成了蛋白质的穿膜部分，每一个亚基提供两个跨膜螺旋。其中一个亚基显示为红色（PDB：2OAR）。

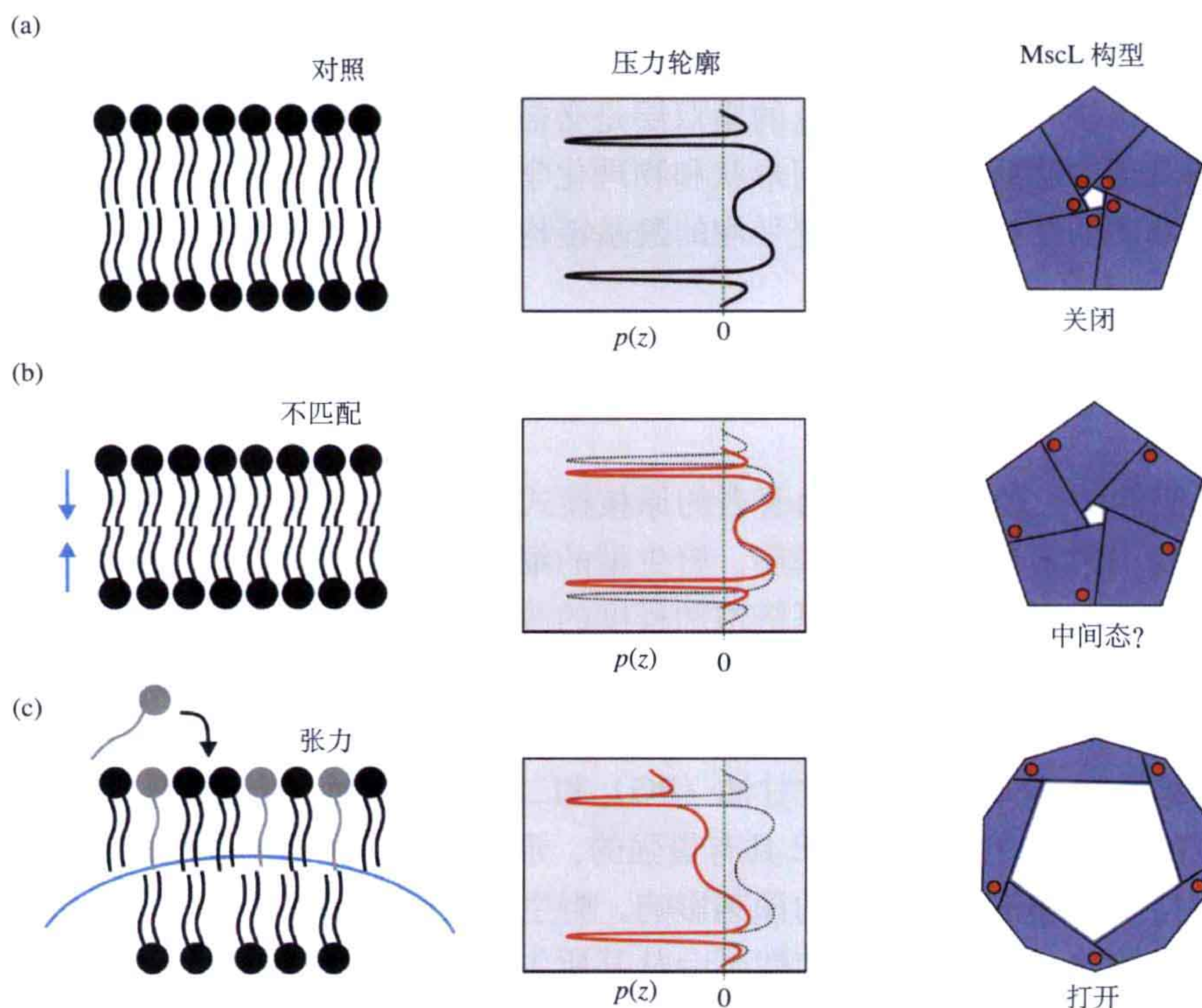


图 4.20 脂双层曲率（或者说横向压力）对于压力敏感蛋白（MscL）的影响示意图。（a）处于静止状态，通道关闭；（b）MscL 位于一个厚度不匹配的脂双层中，压缩或者伸展的压力分布可能激活通道；（c）向膜中添加只有一个烃基链的脂质如溶血磷脂酰胆碱，通道如图（c）所示打开 [经授权根据参考文献 Perozo & Rees. (2003) Structure and mechanism in prokaryotic mechanosensitive channels. Curr Op Struct Biol **13:** 432-442. (Copyright (2003) Elsevier.) 重印]。

4.4 脂质在膜功能中起到基础性作用

生物膜包含多种不同的脂质，许多脂质本身不参与形成脂双层。脂双层对于形成具有功能的、紧密不泄露的细胞膜是必需的，那生物体为何会消耗能量合成不形成脂双层的脂质呢？细胞膜需要在其与所处环境之间建立合适的屏障。细胞中存在可观数量的脂质，它们分开或者扩散在水中时不形成脂双层而是形成其他聚集结构，显然这里有着其他的原因。在过去 20~30 年中，人们对于生物膜中的脂质成分的理解得到了长足的进步。

前面我们看到，脂质形成聚集结构取决于温度、组成、分子结构等一系列参数。下面，我们将会讨论脂质的物理化学性质，生物如何根据细胞所处的环境条件改变其细胞膜中脂质成分，什么决定了膜脂质的横向组织即膜的域形成，以及胆固醇的重要性。

4.4.1 调节膜脂质构成

许多文献证明所有的生物体都会根据其所处的主要环境及生理条件调节其膜脂质构成。这对于保持一个稳定不泄露的脂双层是必需的。细胞可以任意利用下面三种策略来改变脂双层中脂质分子的几何形状和物理化学性质：①改变酰基链的结构；②改变极性头部基团的结构；③不改变平均的酰基链构成而是对其进行重新配置以形成新的脂质类型。

4.4.1.1 大肠杆菌

革兰氏阴性菌大肠杆菌是最为著名的原核模式生物。大肠杆菌只有三种在原核生物和真核生物中常出现的主要膜磷脂。野生型的细胞对于脂质组成的调节来自于对酰基链结构的改变，主要表现在酰基链饱和程度的改变。对于许多生物来说这是一种常见的对于环境温度变化进行的反应。由此，大肠杆菌可以成功地保持其脂质以脂双层的状态存在，而避免了它们形成晶状或者非薄层状液晶相结构。

磷脂酰乙醇胺（PE）、磷脂酰甘油（PG）和二磷脂酰甘油（心磷脂）是野生型大肠杆菌合成的三种主要膜脂质。PE 具有最强的、形成反向非薄层相的倾向，这种特性受到其酰基链长度和不饱和程度的深刻影响。野生型大肠杆菌细胞可以自己合成自身所需的所有脂肪酸，因此这些脂肪酸不会从其所生长的培养基中掺入。当大肠杆菌的生长温度改变时，极性头部基团的组成几乎保持恒定，而其酰基链的饱和性则相应的调整（图 4.21）。通常，温度升高可以使膜脂质的相平衡由薄层状向立方和（或）六角相位（H_{II}）转变。野生型大肠杆菌对生长温度升高的对策是将更短和更饱和的酰基链掺入其膜脂质中。这种改变可以降低 PE 形成非薄层相位的能力，并以此抵消温度升高带来的影响（图 4.21）。

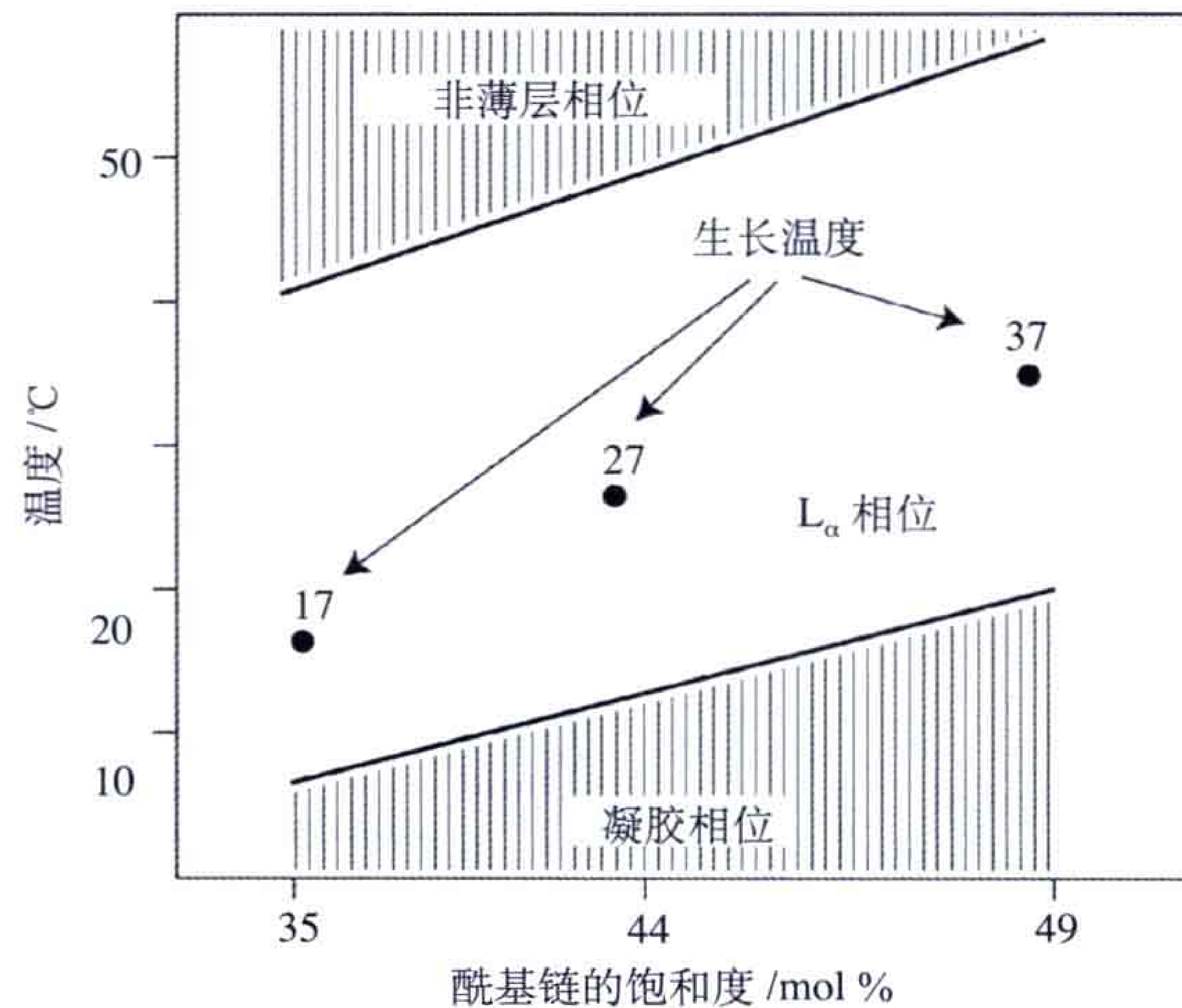


图 4.21　细菌大肠杆菌的生长“窗口”。大肠杆菌需要调整脂肪酰基链的饱和度从而保持无法形成薄层和形成薄层的脂质之间的平衡 [经授权根据参考文献 Morein S, Andersson A-S, Rilfors L, Lindblom G. (1996) Wild-type Escherichia coli cells regulate the membrane lipid composition in a “window” between gel and non-lamellar structures. J Biol Chem 271: 6801-6809. Copyright (1996) HSBMB.) 改编]。

4.4.1.2　莱氏无胆甾原体

缺失细胞壁的细菌莱氏无胆甾原体（*Acholeplasma laidlawii*）A 型菌株对于膜脂质组成的调节得到了详尽的研究。这个物种可以主要通过调节极性头部基团的结构以对某些生长条件加以适应。这样，细胞得以力求保持其脂质在形成脂双层和形成反向非薄层结构之间微妙的平衡关系。因而，基于以下两个原因，这个物种可以成为研究完整细胞膜物理化学性质的绝佳模式生物：①它可以可控地向膜酰基链中引入变化，以及向膜中引入固醇类及其他分子；②从其中可以简单地得到没有污染物的纯净膜成分。早期的差示扫描量热法研究显示，莱氏无胆甾原体膜中的脂质表现出可逆的凝胶 - 液相晶相转换的特性。而且，从 X 射线散射和 ^{2}H 核磁共振波谱的研究可以推断，这些膜是由流动的脂双层结构建造而成。莱氏无胆甾原体对于膜脂质相位平衡的代谢调节与大肠杆菌使用的是不同的策略，但是其对于膜的物理化学性质的影响结果是相同的。

A 型莱氏无胆甾原体可以合成其中的膜脂质（图 4.22）。其中单葡萄糖基双酰甘油（monoglucosyldiacylglycerol，MGlcDAG）、单酰 MGlcDAG（MAMGlcDAG）和单酰双葡萄糖基双酰甘油（MADGlcDAG）三种脂质可以形成非薄层相。磷脂酰甘油（PG）、双葡萄糖双酰甘油（DGlcDAG）、甘油磷酰 -DGlcDAG（GPDGlcDAG）及单酰双甘油磷酰 -DGlcDAG（MABGPDGlcDAG）可以形成薄层状相位。GPDGlcDAG 在高的水含量时也可以自发形成普通的微团状。

莱氏无胆甾原体可以在内源性脂肪酸无法合成的情况下生长，其细胞被迫将外源的脂肪酸掺入其膜脂质中。细胞借由调节极性头部基团的组成以适应掺入的不同酰基链，而其极性头部基团的组成是合乎逻辑的。通常，当酰基链的长度和不饱和性增加时，脂质形成反向非薄层结构的比例降低。对于形成薄层和非薄层相位脂质比例的调节可

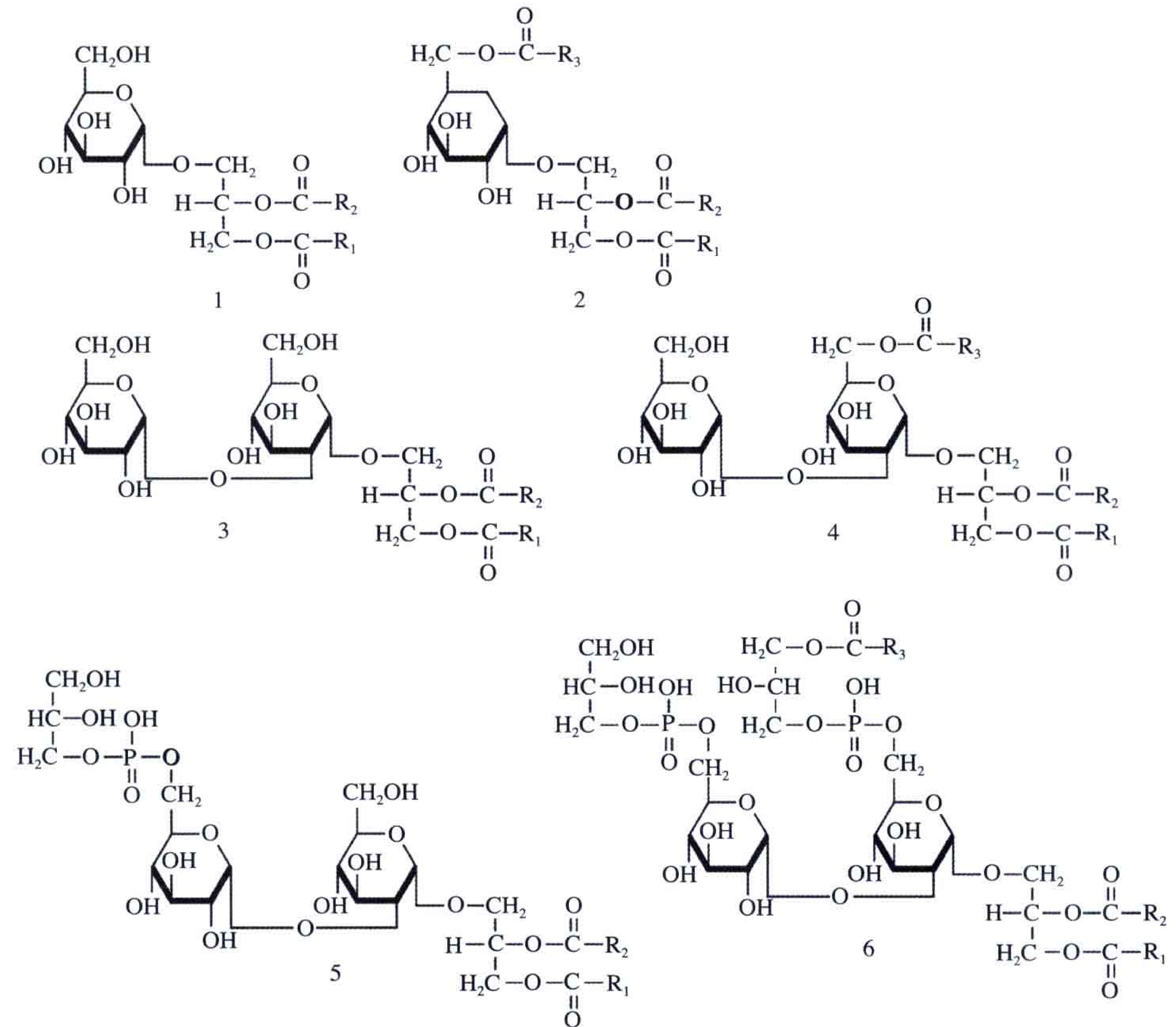


图 4.22 莱氏无胆甾原体膜中糖基化和磷酸化脂质的结构：1. MGlcDAG; 2. MAMGlcDAG; 3. DGlc-DAG; 4. MADGlcDAG; 5. GPDGlcDAG; 6.MABGPDGlcDAG[经授权根据参考文献 Andersson A-S, Rilfors L, Bergqvist M, Persson S, Lindblom G. (1996) New aspects on membrane lipid regulation in Acholeplasma laidlawii A and phase equilibria of monoacyldiglucosyldiacylglycerol. Biochemistry **35:** 11119-11130. Copyright (1996) American Chemical Society. 重印]。

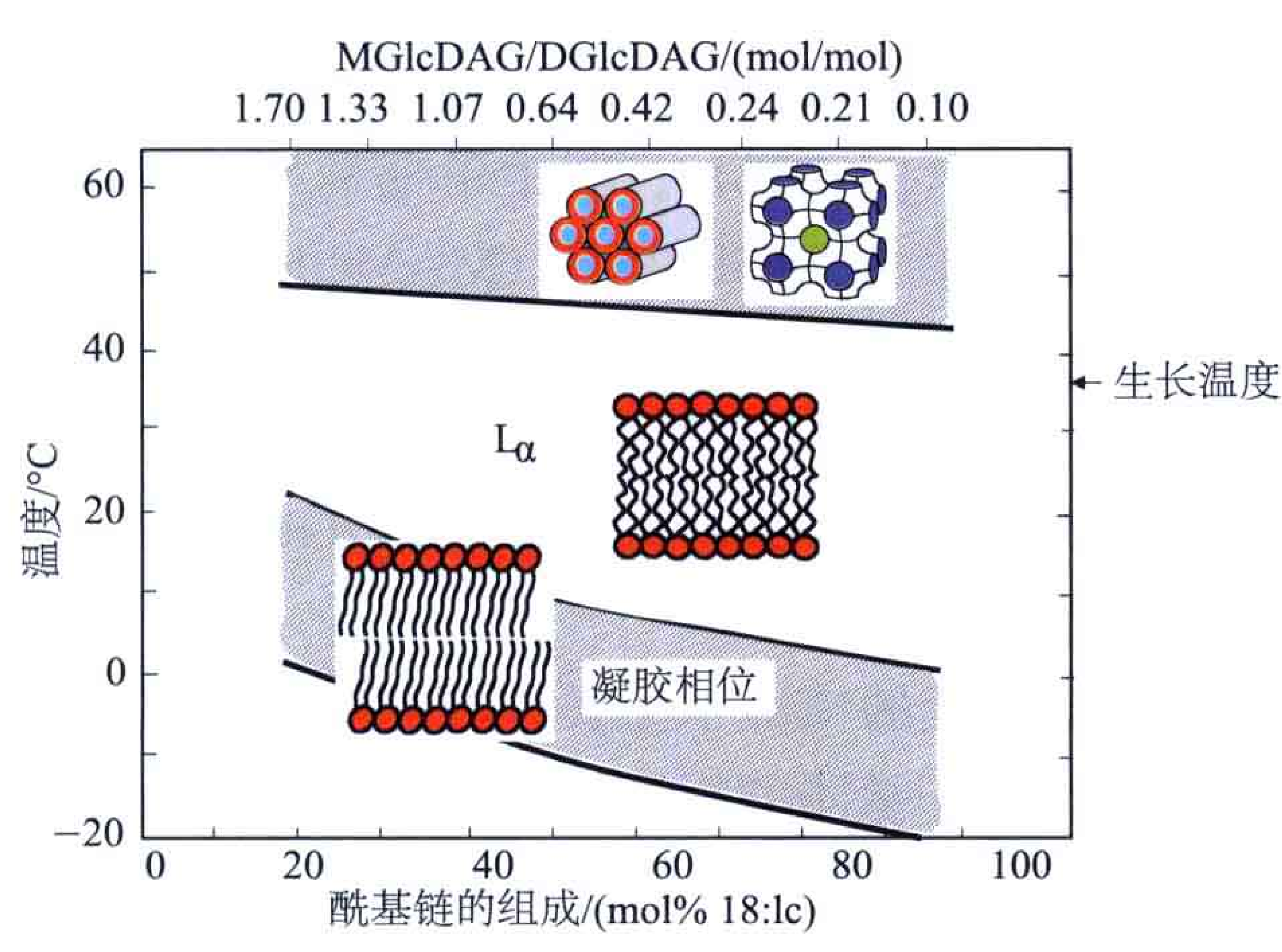


图 4.23 细菌莱氏无胆甾原体在其脂质处于薄层状态时可以生长。图中显示的“窗口”中无法形成薄层和形成薄层的脂质形成平衡。注意上面的相位向非薄层状态转变的温度几乎是恒定的。细菌生长的温度比这个转变温度低10~15℃。膜中脂质的组装可能至关重要 [经授权根据参考文献 Lindblom G, Rilfors L. (1989) Cubic phases and isotropic structures formed by membrane lipids — possible biological relevance. Biochim Biophys Acta **988:** 221-256. Copyright (1989) Elsevier. 改编]。

以在非常窄的间隙内对总脂质提取物产生不同的相变温度（图 4.23）。还要注意的是酰基链的不饱和度和长度对于细菌的生存有着重要的作用（图 4.24）。为了保持形成薄层和非薄层脂质之间的严格平衡，在短的饱和酰基链（图 4.24 中蓝色区域的左手部分）存在时，细菌甚至要合成以前未见的新的脂质，这说明膜中脂质的组装对其功能是非常重要的。

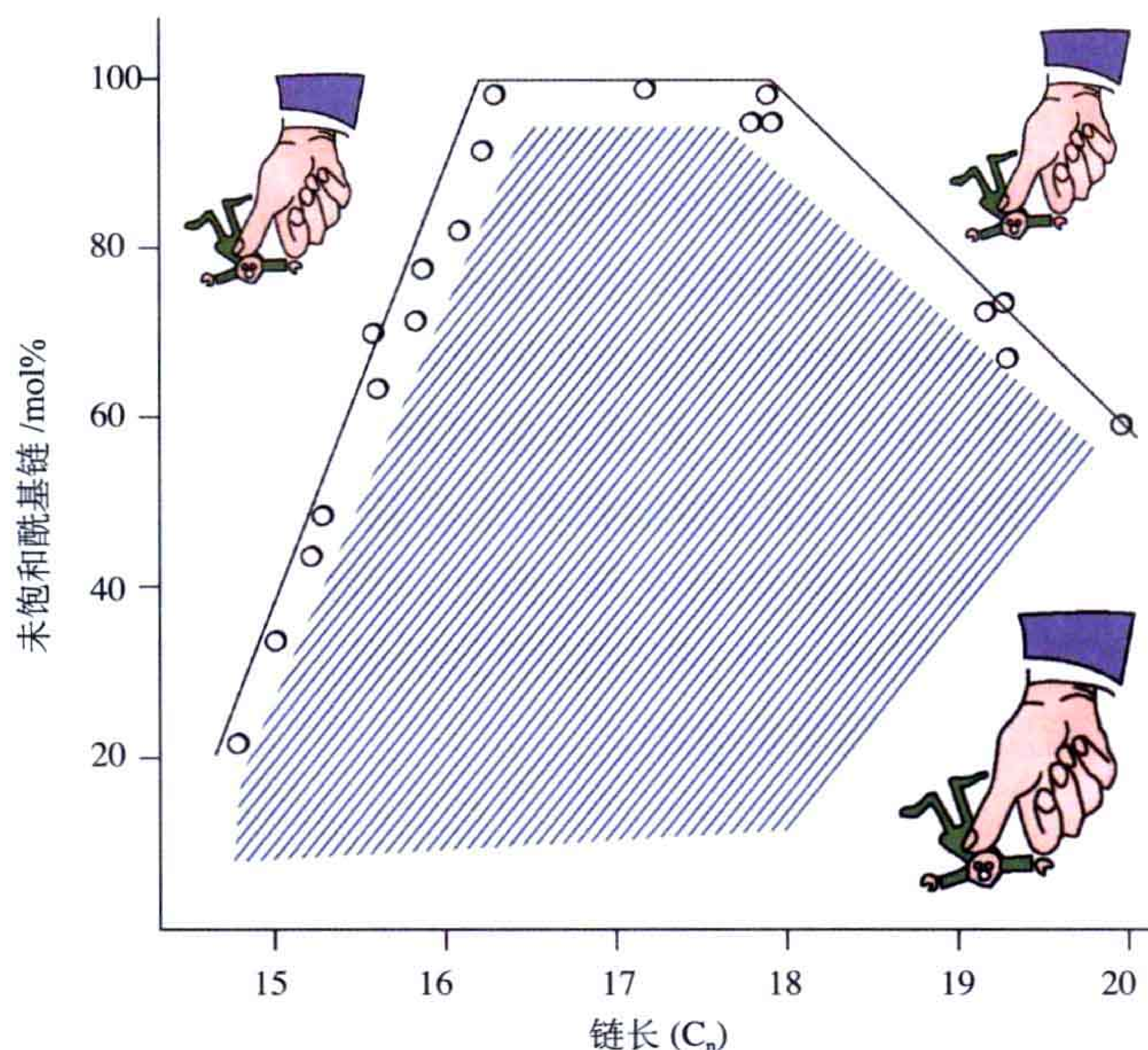


图 4.24　存活的莱氏无胆甾原体膜中脂质酰基链的不饱和度是链长度的函数。莱氏无胆甾原体仅能在图中蓝色阴影区域内存活，其脂质的脂肪链长度和不饱和性处于最优的组装关系，无法形成薄层和形成薄层的脂质赋予质膜合适的物理化学性质。在蓝色阴影区域的左手边，即酰基链非常短的时候，细菌需要合成新的具有三个酰基链的脂质 MADGlcDAG（图 4.22）以保持生存。在细菌具有稍长的 16~17 个碳的低不饱和酰基链时，细菌合成另一种脂质 MAMGlcDAG[经授权根据参考文献 Andersson A-S, Rilfors L, Bergqvist M, Persson S, Lindblom G. (1996) New aspects on membrane lipid regulation in Acholeplasma laidlawii A and phase equilibria of monoacyldiglucosyldiacylglycerol. Biochemistry **35:** 11119-11130. Copyright (1996) American Chemical Society. 改编]。

4.4.1.3　真核生物

在研究莱氏无胆甾原体和大肠杆菌的过程中可以得到一个重要的结论，即细胞总是会调节膜脂质的组成以保持一种薄层状的液晶相位，从而避免形成凝胶或者非薄层状液晶相位（图 4.23 和图 4.21）。在其他原核生物中，膜脂质组成和脂质的物理化学性质之间的关系也得到研究。酪酸梭状芽孢杆菌（*Clostridium butyricum*）和巨大芽孢杆菌（*Bacillus megaterium*）可以与莱氏无胆甾原体及大肠杆菌使用同样的方式调节其膜脂质组成。同样，真核生物如真菌、高等植物和冷血动物（体温可以随环境变化的动物）经常会根据其周围温度以改变它们的膜脂质组成。这可能暗示着对温度的适应性可以改变脂质的相变温度，从而使流动的双层结构保持稳定，而周围的温度被凝胶相和非薄层相限制在一个特定的范围内。

在植物中也有与这类似的对膜脂质组成的调节作用。燕麦和黑麦叶片的质膜对冰冻的耐受性也主要是冰冻引发的脱水从而使膜不稳定的结果之一。从春天生长的燕麦叶片与冬天生长的黑麦叶片提取的质膜中，脂质组成有极大不同。春天的燕麦质膜中含有大量的磷脂类、脑苷脂类和酰基化的甾醇糖苷类物质；而冬天的黑麦中则含有更多的磷脂、更少的脑苷脂及大量的自由甾醇类物质。然而，在两种物种中，冷适应都带来磷脂类含量的升高、脑苷脂组分的下降。研究显示在冰冻创伤时脂质会形成 H_{II} 相位。H_{II} 相位的发生率与原生质体和叶片组织的致死性创伤紧密相关，这可以由原生质体对渗透压反应性的丧失及叶片细胞内容物的泄露显示出来。冬黑麦和春燕麦之间冰冻诱导形成 H_{II} 相位所依赖的起始温度具有显著不同，它与质膜脂质组成的不同是相关的，这并不奇怪。然而，当两种植物对冷适应后，由于质膜中的脑苷脂成分大幅下降，则不会产生冷冻诱导的 H_{II} 相位形成。考虑到脑苷脂在较低温度下会促进 H_{II} 相位的形成，这种适应性就不难理解了。脑苷脂的极性头部基团的亲水性较低，从而使其分子形状更像楔形，而这使其更容易包装形成 H_{II} 相位。因此，黑麦和燕麦的质膜之所以会有在冷冻诱导时脱水而从薄层相向 H_{II} 相位转变的特点，应该是膜脂质物理化学性质包括双层膜表面水化及脂质的组装等造成的。

4.4.2　不能形成薄层状的脂质在膜功能中的作用

薄层状液晶相位中的多层脂双层结构已经被用作生物膜的模型很长时间，而其中的单层脂双层小泡则在制药业中有广泛的应用。对于形成非薄层结构的了解使得人们对于膜脂质在细胞过程中所起的作用逐渐有了新的认识。越来越多的实验和理论证据显示，脂质活跃地参与了细胞的许多重要功能。令人惊奇的是，在以前的生物化学教科书中很少提到关于脂质的这一点。

4.4.2.1　特殊的膜结构

细胞合成不能形成薄层结构的膜脂质并使其与可以形成薄层结构的膜脂质保持一定的平衡，其中的一个原因是细胞需要这些无法形成薄层结构的脂质形成非脂双层结构或者具有小曲度半径的脂双层结构。具有小曲度半径的脂双层结构在多种生物膜中都有出现，包括内质网、线粒体内膜及叶绿体类囊体膜的基粒层叠（图 4.2）。特别是光面内质网和线粒体内膜的电子传递显微镜图片显示它们与双连续性立方或者 L_3 相位结构相似（图 4.15）。白色质体是于生长在黑暗中的植物叶片中发现的一种细胞器，其中存在一种被称为原片层体的高度规则的、分支的管状膜结构。在经过光照之后，白色质体转变为叶绿体，而原片层体则发育成为线粒体的类囊体膜。

4.4.2.2　对于膜结合蛋白活性的影响

20 世纪 80 年代的研究表明，当脂质形成非双层结构，或者脂双层结构的膜中掺入

会使其不稳定的分子时，膜蛋白在膜重构过程中的插入效率及膜蛋白的活性都会增强。蜡状芽孢杆菌（*Bacillus cereus*）的卵磷脂特异性磷酯酶 C 的活性在使薄层相位不稳定的脂质分子存在时增强。这种增强作用来自“挫折的”脂双层的组装压力，而不是实际形成的反向非薄层相位。这种“挫折的”脂双层对于锚定和激活膜周边蛋白有一定的作用。对蛋白激酶 C 来说，酶掺入膜中的比例及膜结合形式的酶的活性，在存在不能形成薄层结构的脂质时都会升高。真菌中的多肽丙甲菌素（阿拉霉素）在膜中会形成一种电压依赖性离子通道，这个离子通道在脂双层中无法形成薄层的脂质比例升高，即两个脂单层的自发曲度升高时，会更容易呈现一种高传导性状态。

4.4.3　膜融合

膜融合是在所有细胞中都非常重要的现象，它与膜运输、小泡介导的运输、精子 - 卵子融合及病毒细胞融合等过程有关。在过去的 10 年中，对于融合过程的分子机制在实验上和理论上都做了详尽的研究。这里我们仅讨论发生在简单的脂质系统中的融合现象，以及有非薄层结构参与融合过程的特定系统中的融合现象。融合的关键步骤在于将两个相对的膜中的脂质分子重组形成一个连续的膜。为了理解融合的过程，我们需要考虑动力学和结构两个方面。

早先有人提出，为了在膜融合过程中重组膜结构，需要破坏脂双层以形成其他的聚集结构。溶血磷脂胆碱（LPC）可以介导红细胞之间的，以及红细胞和成纤维细胞之间的融合，而人们已知 LPC 在高水含量时会形成普通的微团溶液（参考图 4.12）。在添加合适的试剂使完整的红细胞融合时，会形成反向的非脂双层聚集结构。最近，有人提出了膜融合的一种修正的茎环结构（由薄层向反向非薄层相位转化时出现的结构）机制。大量的理论计算和模型构建显示在薄层相位向 H_{II} 相位或者立方相位转化时也会存在类似的中间态结构。这些结果意味着在模型膜或者在生物膜中，低水平的特定脂质（如易溶性脂质 lysolipid)，或者一些重量融合多肽或融合蛋白的存在可以增强融合效率（图 4.25）。

脂质膜融合的第一步是在接触的膜外层之间形成一种茎环状的连接，而膜内层和内部液体内容物仍然保持完整。形成茎环结构后下一步是将展开的脂质融合孔隙打开以完成融合反应。

脂质膜的融合仅在特定的脂质构成及周围水浴溶液中存在特定的离子或者在膜内接触并脱水的条件下才能发生。在初始状态下，膜脂单层积聚能量并在融合过程中释放。如果相接触的膜单层的曲率与它们的自发曲率不一致，促进融合的能量就会升高。

蛋白质可以影响接触的膜中脂质的组成，从而使脂质具有负的自发曲率。因而，可以起始酶促级联反应使特定脂质如双酰基甘油和 PE 浓度升高的磷脂酶和酰基转移酶可以促进融合，并确实参与到一些细胞内的融合反应过程中。融合蛋白可以使脂双层结构扭曲，从而引发弹性压力并导致融合。例如，感冒病毒的红细胞凝集素（HA）介导的融合完全取决于插入膜的融合多肽结构域形成的一种特殊的回旋镖样构象，这种构象被认为可以形成融合所需要的脂双层的扭曲（见 15.2.1 节）。

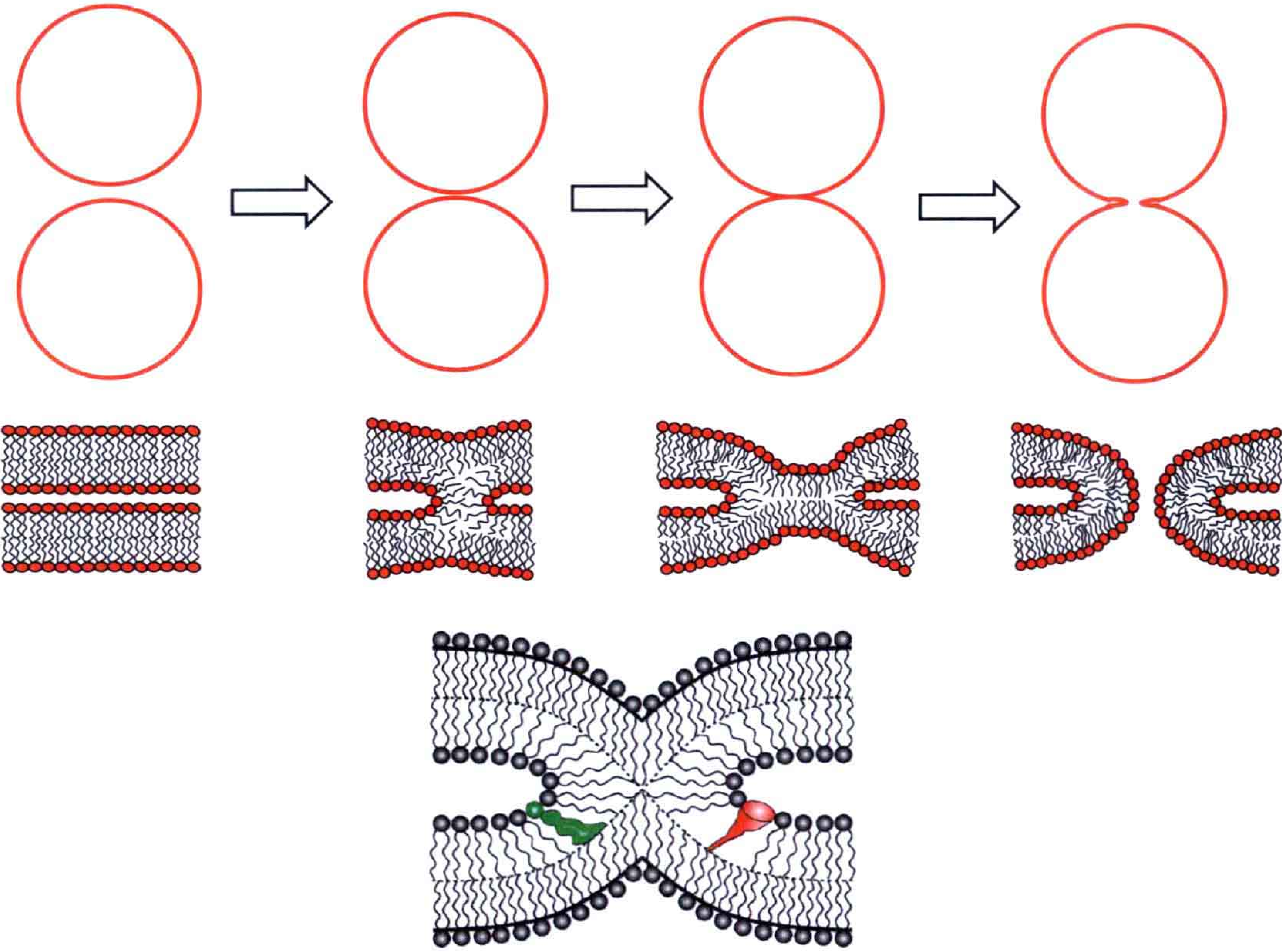


图 4.25 上，膜融合的步骤。开始和结束阶段在两端，中间两个是中间状态。第三阶段中，茎环形成。脂质的形状对于形成茎环非常重要（参照脂双层曲率的章节）。下：膜融合过程中出现的茎环结构假想图。虚线表示脂双层中腔。绿色圆锥形脂质分子如 PE（茎环的左侧翼部）与茎环结构相配，而红色的倒圆锥形脂质分子如 LPC（茎环的右侧翼部）会破坏茎环脂质的组装。

起始融合中间体，即融合茎环的能量大约为每摩尔几万卡路里，其特征区域面积为 100nm^2。因而，驱动这个融合阶段所需要的压力（单位面积的能量）至少为 0.1 kcal/mol/nm^2。

4.4.4 合成脂质的酶

脂肪酸合成酶的结构描述见 5.4 章。膜脂质组成的调节作用暗示着合成脂质的酶（脂合成酶）的活性会根据细胞的主要生长环境加以调节。因而，一些反映脂双层状态的信号必须从脂双层被传递到脂合成酶中。通常来说，脂合成酶或多或少都与脂双层紧密相连，因此一种可能性是这些酶的活性直接受到脂双层特征的影响（图 4.26）。另一种可能则好似合成活性受到与酶结合的效应分子的调节。这些效应分子包括膜脂质，或者依次感受脂双层状态的特殊蛋白质。

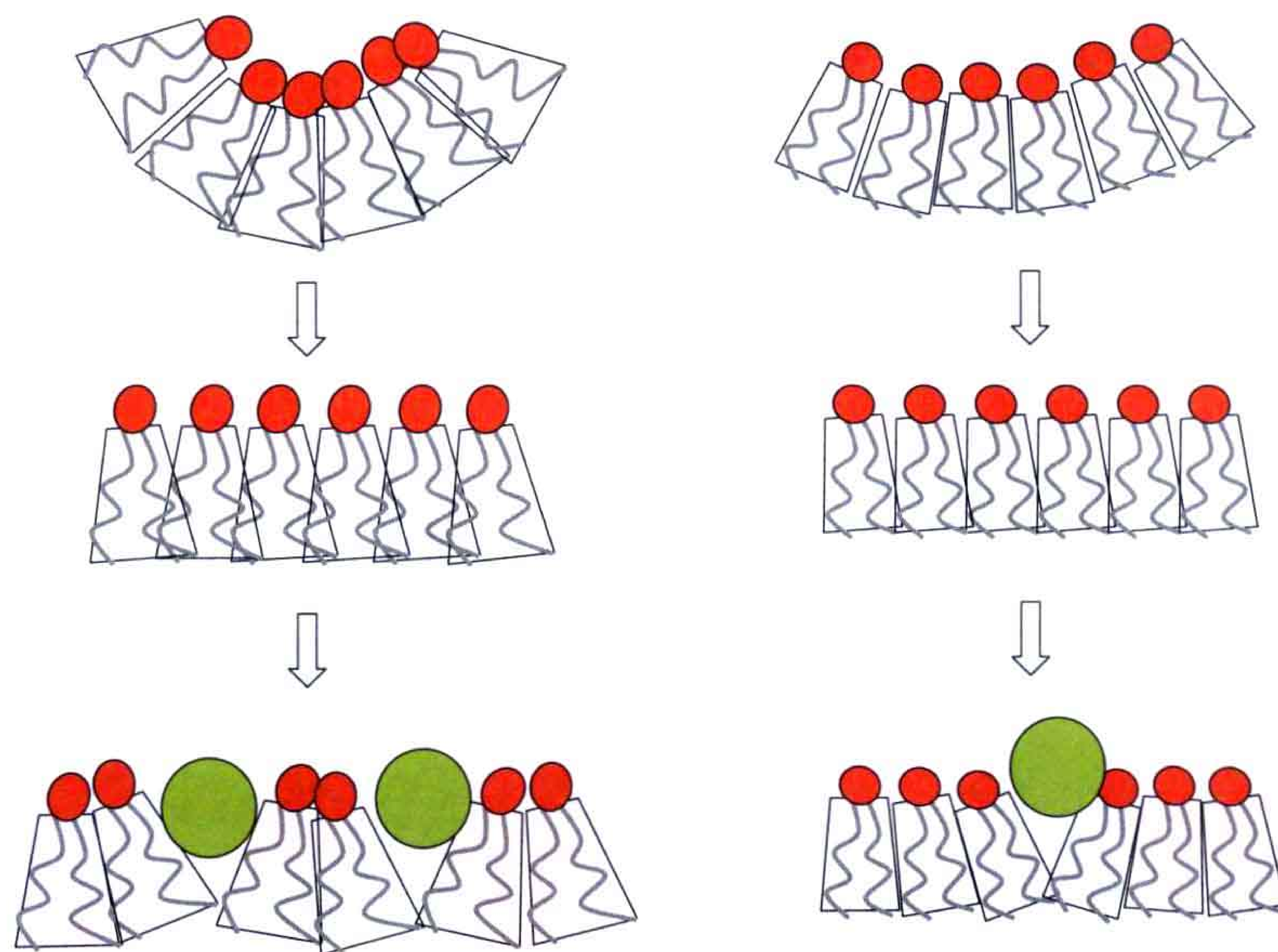


图 4.26　细胞通过将 CCT（绿色）与 DOPE（左）或者 DOPC（右）脂双层结合以调节膜弹性张力的卡通示意图。由于两种脂质分子的形状不同，使 DOPE 脂双层松弛（平坦）比使 DOPC 脂双层松弛需要更多的 CCT 分子 [经授权根据参考文献 Attard GS, Templer SH, Smith WS, Hunt AN, Jackowski S. (2000) Modulation of CTP: phosphocholine cytidylytransferase by membrane curvature elastic stress. Proc Natl Acad Sci USA **97:** 9032-9036. 改编]。

4.4.4.1　三磷酸胞嘧啶核苷（CTP）：磷脂酰胆碱

磷脂酰胆碱（PC）是多数真核细胞的一种主要膜脂质。因而，对于 CTP 而言，磷脂酰胆碱胞苷酰转移酶（CCT）活性的调节对于膜生物非常重要。CCT 的活性在阴离子磷酸酯和中性脂类如双酰基甘油存在的条件下升高。CCT 的一段两亲 α 螺旋肽段可以结合阴离子脂质小泡，阴离子脂质的激活作用来自于这些脂质和两亲螺旋中的碱性氨基酸残基的静电相互作用。

CCT 的活性受到脂质膜单层膜中存储的曲率或者弹性压力的调节。当脂单层中的弹性压力上升时，它也单调递增（图 4.26）。相反，当少量的去垢剂分子掺入脂双层，即脂单层的自发曲率下降时，酶活性显著降低。这样，一个纯粹的物理反馈信号可以在调控脂质膜的合成中发挥关键的作用。

同样的细胞模型可以用于描述质膜脂质以及 CCT 的生物合成，它们可以将莱氏无胆甾原体的膜结合酶的生物化学与生物物理性质相结合。这个模型意味着脂双层中储存的弹性能量可以通过与两亲性的α螺旋的相互作用来改变一些曲率敏感的酶的活性。由于它们的结合取决于脂质的构成，所以其存储的弹性能量的调节取决于脂双层中脂质的组装，这是一个生物物理的反馈机制。质膜中形成薄层的和不能形成薄层的脂质必须保持平衡，而特定脂质的浓度也必须加以限制。利用莱氏无胆甾原体中测量得到的脂质曲率的值，虽然还不能做到完全定量，理论模型对于膜过程的描述已经非常好。

4.4.5 膜中的脂质区域和膜木筏

1972 年，Singer 和 Nicolson 提出了一个膜的经典模型，他们认为膜是一种基质，蛋白质在脂质的“海洋”中进行一个自由度的运动。这种“液态镶嵌模型”已经成为我们现在理解膜双层及其生理机能的基本架构和基准（图 4.27）。

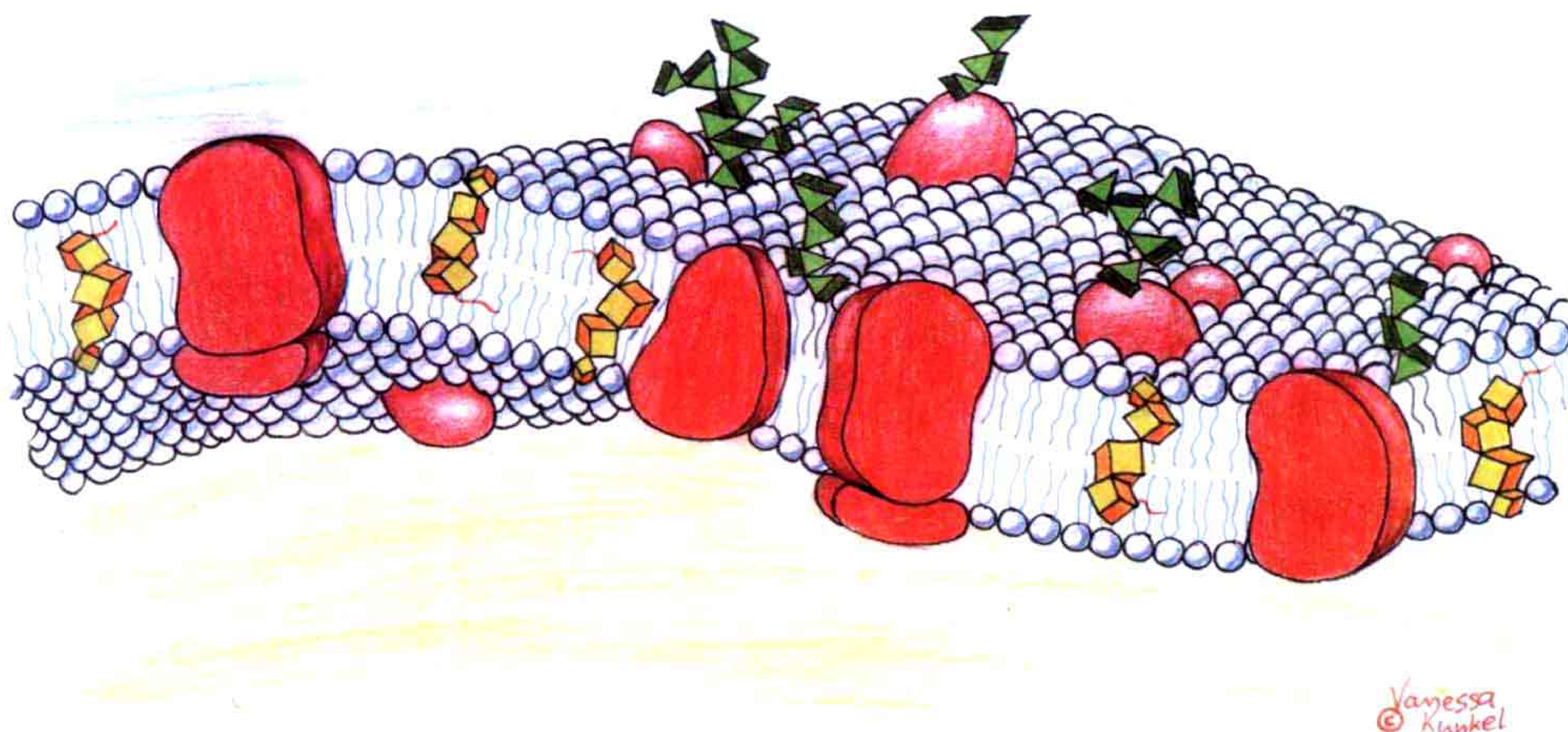


图 4.27 1972 年 Singer 和 Nicolson 提出的生物膜流体镶嵌模型示意图。黄色的跨膜分子是胆固醇，绿色的伸向溶液中的部分代表糖分子（Vanessa Kunkel 惠赠）。

然而，这个模型认为膜具有均一性，表现在组成膜的分子在膜中是随机分布的，这个观点后来被修正。近来许多研究表明细胞膜具有非常复杂的横向组织。例如，利用单颗粒跟踪技术发现，被标记的脂质或者蛋白质分子在横向表现为扩散性运动，并被暂时限制在膜中的一个范围之内。

膜中富含胆固醇、鞘脂类和甘油磷脂的横向区域（被称为脂质木筏）是膜生物学研究的中心课题（图 4.28）。这些液相膜中的纳米级别大小的木筏区域被认为在信号转导、脂质运输、调节膜蛋白的活性及胞转作用（大分子经囊泡由细胞的一侧向另一侧运输）等方面起到支配性的作用。

脂质区域的形成首先在能够承受去垢剂 TritonX-100 的溶解作用的去垢剂抗性膜（DRM）中被观察到。DRM 富含胆固醇、总体饱和的鞘磷脂，而在膜的其他区域则包含更多的非饱和脂质，主要是卵磷脂。不同的细胞膜中含有木筏结构的证据稳定增加，虽然大多数证据不是直接的。因而，生物膜中存在稳定的木筏区域的说法仍然存在着研究和质疑。其中一个问题是膜中的木筏结构可能太小而无法被荧光显微镜技术观察到，而在类似的研究中荧光显微镜是最为常见的研究技术。因此进行直接观测活细胞中膜木筏结构如其大小、寿命、动态及脂质组成的实验是非常困难的。另一方面，对脂质模型膜的研究则能够提供非常多的信息，现在对于脂质区域的存在或者膜中横向

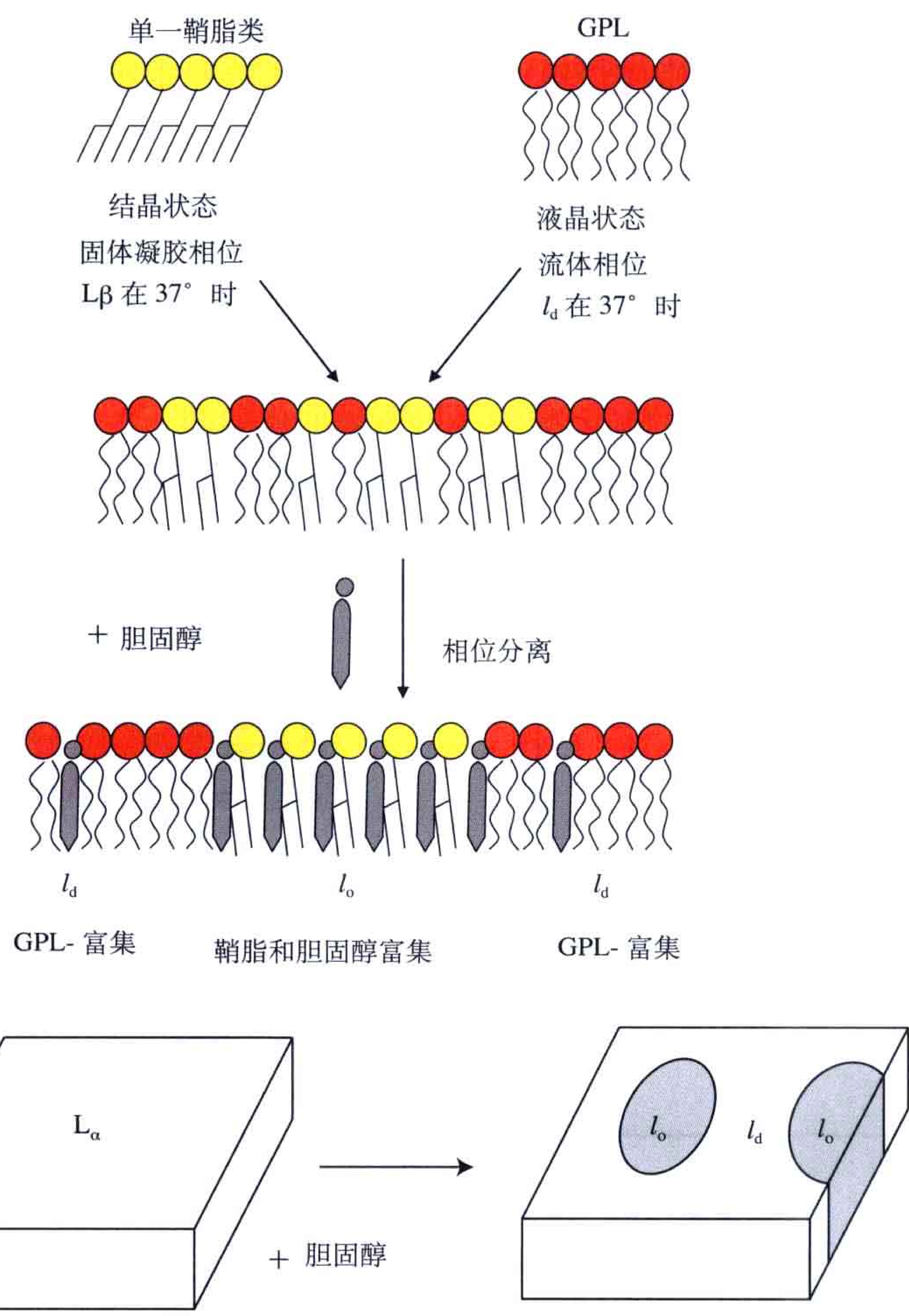


图 4.28　上：胆固醇诱导含有甘油磷酸脂和鞘脂的膜产生横向相位分离形成不同的区域。GPL 指甘油磷酸脂。在下部膜中的黄 - 灰色部分代表木筏结构。下：流动脂双层添加胆固醇而形成木筏区域的示意图。注意光滑圆形截面产生的原因是脂双层中的 l_o 区域同样是液相的。L_α 代表薄层相位；l_o 是规则的脂质区域。

相位分离的说法已经被广泛接受。

使用原子力显微镜、荧光显微镜、荧光猝灭、单颗粒示踪、差示扫描量热法、固相和核磁共振扩散法、X 射线衍射等方法，可以观察到脂质区域结构在多层脂双层系统、脂质单层，或者大的脂质小泡中也存在。图 4.29 是通过这些研究所绘制的一个相图的例子。一张详细的相图对于理解脂双层结构中出现脂质区域的驱动力是非常重要的。实际上，迄今为止多数相关的脂质系统的相行为都与最初发表的双软脂酰磷脂酰胆碱（DPPC）、胆固醇（CHOL）和水的三元系统相似。在这种相图中，人们发现液态 − 规则相（l_o）和液态 − 不规则相（l_d）彼此平衡。这两种相位都处于流动液晶状态，但是在 l_o 相位中的烃基链比 l_d 相位更加规则或者说更加延展。相图中 l_o 和 l_d 相位之间有一大块双相区域。细胞膜中的木筏结构被认为是处于这个区域的 l_o 相位结构。

木筏结构被认为含有高水平的胆固醇、鞘脂及饱和的磷脂。鞘磷脂或者鞘糖脂与胆固醇共同存在可以促进脂质烃链的规则化，这意味着木筏结构具有与 l_o 相位类似的结构。

细胞膜穴样内陷（胞膜窖）是脂质木筏的一个亚型（图 4.30）。胞膜窖是内皮细胞、脂肪细胞和平滑肌细胞膜表面微型的烧瓶状膜内陷区域。胞膜窖中主要的蛋白质是窖蛋白（caveolae），一种可以有效结合胆固醇并能结合多种信号大分子包括 G 蛋白、钙调蛋白等的脚手架蛋白。窖蛋白也可以调节细胞内和细胞表面的胆固醇水平。对缺失窖蛋白 -1 从而没有胞膜窖的基因敲除小鼠所做的实验显示，此小鼠在动脉舒张、肌原性张力、运动耐量等方面都有显著缺陷，其原因是细胞信号和 NO 代谢的不正常。

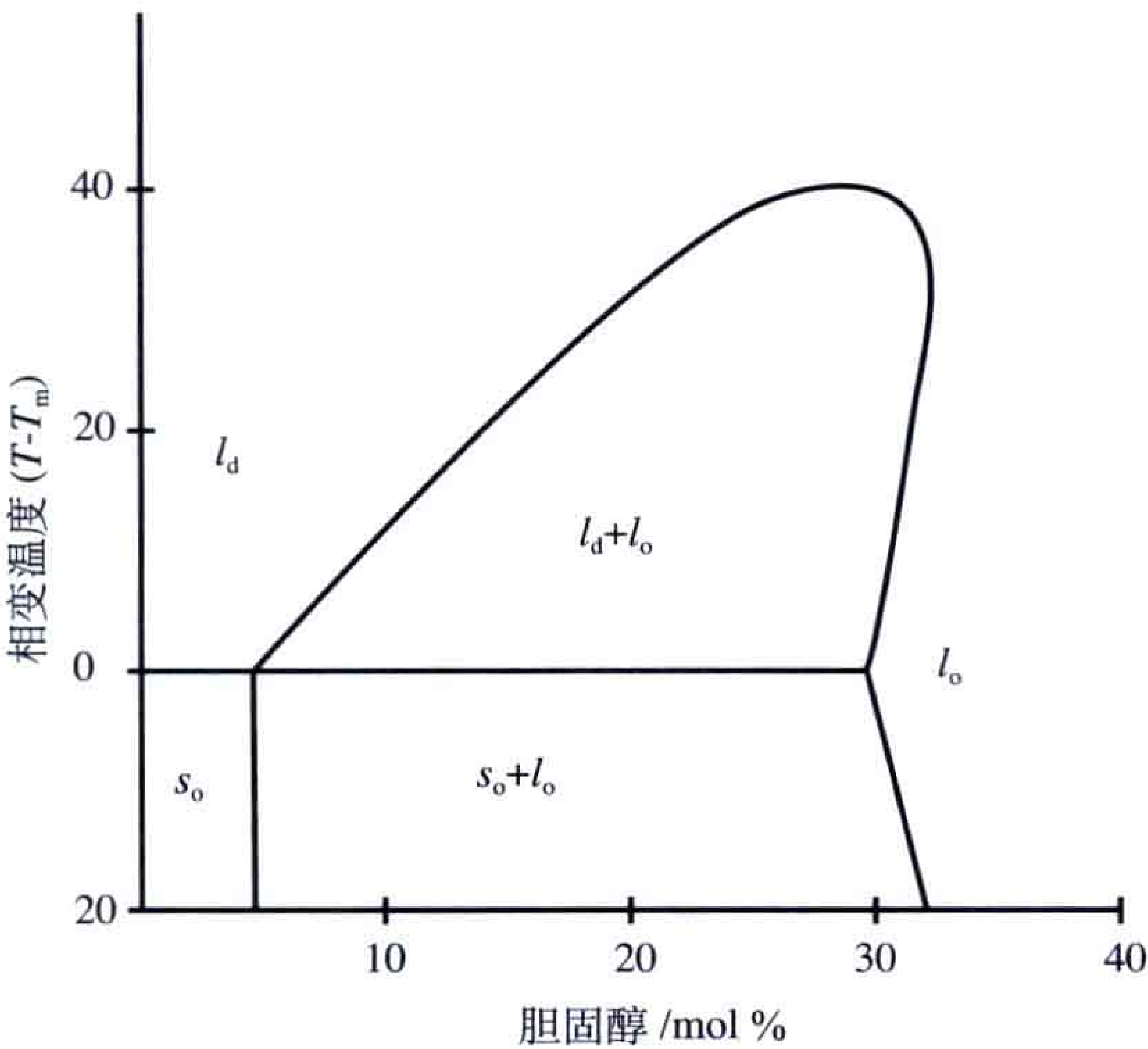


图 4.29　饱和磷酸酯、胆固醇（CHOL）和水组成系统的通用温度 / 浓度相图。注意混合相区域包含不规则的 l_d 和规则的 l_o 相位。s_o 代表凝胶相位；T_m 是有凝胶向薄层相位发生相变的温度。

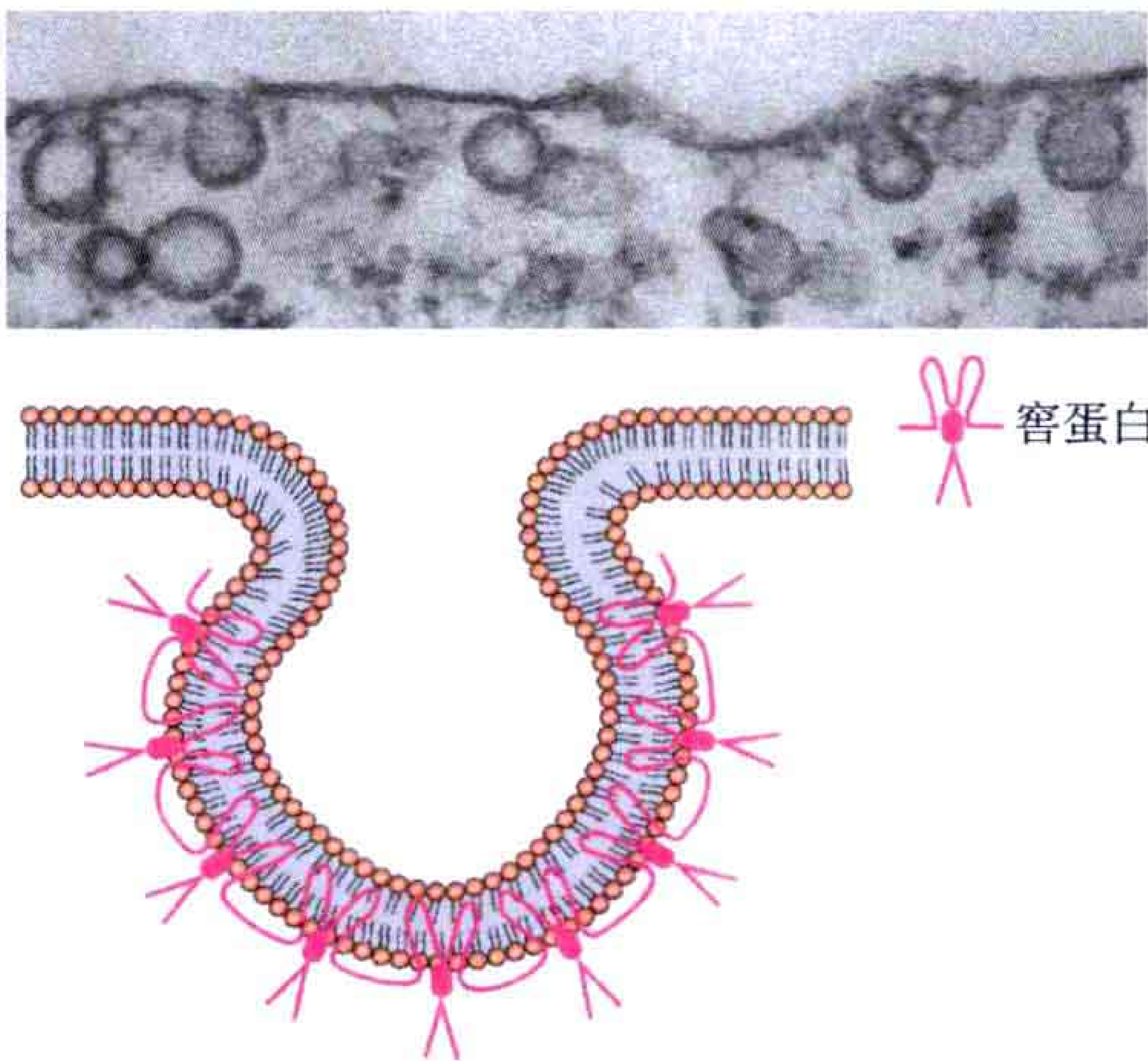


图 4.30　质膜上被称为窖的瓶装凹陷图。

脂质木筏被认为在多种疾病中起作用，包括动脉粥样硬化、高血压、阿尔茨海默病、朊蛋白病及病毒感染等。

我们如何才能理解脂双层中区域形成的驱动力呢？最近，一些以前没有得到重视的脂双层物理化学特征给我们提供了新的有用信息。因此，横向的相位分离形成 l_d 和 l_o 相位被认为是熵变驱动的。此熵变来源于将不饱和的脂质掺入主要由饱和脂质和胆固醇组成的高度规则的相位时的困难性。不饱和的脂质倾向于形成 l_d 相位，而饱和的脂质则倾向于形成 l_o 相位。

这一章中，我们着重描述了相行为和膜脂质，特别是无法形成薄层的脂质在膜多种功能上的参与。越来越多的证据证明许多膜脂质分子可以作为细胞的信号分子起到重要作用。与已经建立的功能基因组学和功能蛋白组学类似的，这些研究带来了一个新的学科，即功能脂质组学。

延伸阅读

系统命名法及网址

Fahy E, *et al*. (2005) Acomprehensive classification system for lipids. *J Lipid Res* **46:** 839-861.

LIPID MAPS http://www. lipidmaps. org; http://lipidlibrary. co. uk; http://lipidbank. jp; http://www. lipidat. chemistry. ohio-state. edu; and http://www. cyberlipid. org

International Union of Pure and Applied Chemists and the International Union of Biochemistry and Molecular Biology (IUPAC-IUBMB) (see further reading for URL address http://www. chem. qmul. ac. uk/iupac/

Christie WW. What is a lipid? http://www.lipidlibrary. co. uk

原始文献

Alley SH, Ces O, Templer RH, Barahona M. (2008) Biophysical regulation of lipid biosynthesis in the plasma membrane. *Biophys J* **94:** 2938-2954.

Morein S, Andersson A-S, Rilfors L, Lindblom G. (1996) Wild-type *Escherichia coli* cells regulate the membrane lipid composition in a “window” between gel and non-lamellar structures. *J Biol Chem* **271:** 6801-6809.

Rilfors L, Lindblom G. (2002) Regulation of lipid composition in biological membranes-biophysical studies of lipids and lipid synthesizing enzymes. *Colloids Surf B-Biointerfaces* **26:** 112-124.

综述文章

Chernomordik LV, Zimmerberg J, Kozlov MM. (2006) Membranes of the world unite! *J Cell Biol* **175:** 201-207.

McIntosh TJ (ed). (2007). Lipid rafts. *Methods Mol Biol* 398. Springer Verlag, Berlin.

Mouritsen OG. (2005) *Life-as a Matter of Fat. The Emerging Science of Lipidomics*. Springer

Verlag, Berlin/Heidelberg GmbH & Co. K.

Yeagle PL. (2005) *The Structure of Biological Membranes*. CRC Press, Boca Raton, Florida, USA.

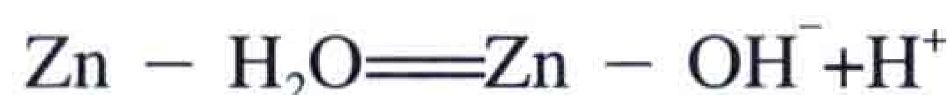
（王开拓 译　金坚石 校）

第五章
酶

酶 (enzyme) 是指可以催化生物学反应的蛋白质。它们催化的反应各式各样，通常可以按照它们催化的反应类型进行分类，如水解酶 (hydrolase)、连接酶 (ligase)、还原酶 (reductase)、氧化酶 (oxidase) 等。在本章中我们将讲述 5 种酶，它们分别在生物体内扮演不同的角色并具有不同的催化模式，其中第一种酶催化一个非常快速的反应，第二种酶在生物体内受到严格的调控，第三种酶是一个分子马达，第四种酶是一个分子开关，第五种酶则是一个具有多重功能的酶。更多类型的酶会在本书其他章节继续讨论。

5.1 碳酸酐酶——一个非常快速的酶

碳酸酐酶 (carbonic anhydrase) 催化一个非常简单的反应，即二氧化碳的水合反应及其逆反应。碳酸酐酶广泛存在于所有已知的细胞中，有 4 种主要类型，即 α 型、β 型、γ 型和 δ 型。碳酸酐酶中有许多同工酶 (isoenzyme) 分别执行不同的生理学任务，这种情况在 α 型碳酸酐酶中尤其多见。4 种类型的碳酸酐酶在演化上有着相互独立的起源，并且也有着不同的三级、四级结构，但却以一种十分相似的方式催化反应（图 5.1）。这个酶最让人好奇的一点是它的催化速率。例如，人类的 α 型碳酸酐酶同工酶 II 在催化二氧化碳水合时的最大转换数（maximal turnover rate）是 $10^6 s^{-1}$，这已经非常接近扩散极限。究竟碳酸酐酶有着怎样的机制使得它可以如此快地行使功能？先来看看它所催化的反应步骤：



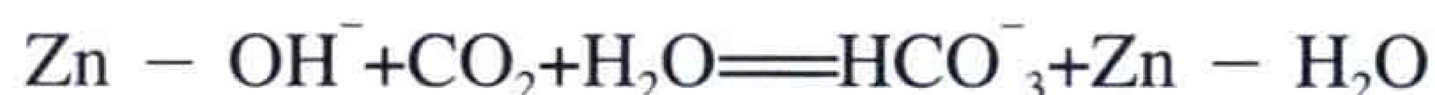
$$Zn - H_2O \rightleftharpoons Zn - OH^- + H^+$$



$$H^+ + His \rightleftharpoons His\text{-}H^+$$



$$His - H^+ + Buffer \rightleftharpoons His + Buffer - H^+$$



$$Zn - OH^- + CO_2 + H_2O \rightleftharpoons HCO_3^- + Zn - H_2O$$

在碳酸酐酶的活性位点中一般都存在着一个锌离子。在 α 型和 γ 型的碳酸酐酶中，这个锌离子结合在三个组氨酸残基上；而在 β 型碳酸酐酶中则是一个组氨酸残基和两个半胱氨酸残基。 于是在 α 型和 γ 型的碳酸酐酶中这个锌离子的电荷并未被其蛋白质配体中和，其所带电荷为 2+（图 5.2）。

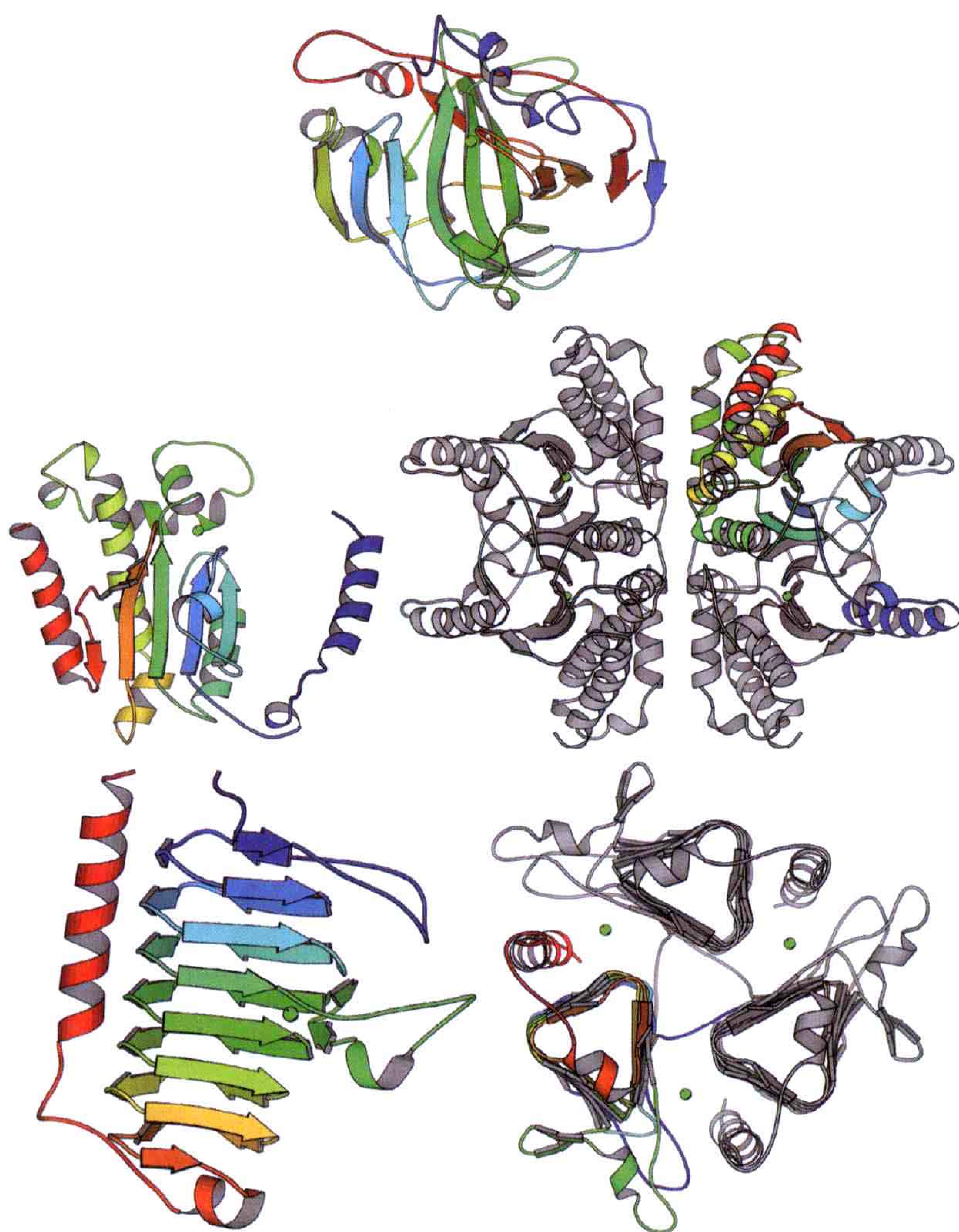


图 5.1　三种不同类型碳酸酐酶的结构图，从上至下分别为 α 型（PDB：2CBB）、β 型（PDB：1I6P）及 γ 型（PDB：1QRE）。在所示的三个结构中可以看出，它们的折叠方式是完全不同的。另外，α 型的碳酸酐酶是单体，而 β 型是由两个或者 4 个二聚体组成的四聚体或者八聚体，γ 型碳酸酐酶则是由 β 折叠组成的三聚体。在图中所示的 β 型和 γ 型碳酸酐酶聚体结构中（第二行和第三行右侧），都用彩色标出了其中的一个单体。通过锌离子的位置（图中绿色圆点所示）可以得知图中每一个碳酸酐酶活性位点的位置。并可见在 γ 型碳酸酐酶中，其活性位点位于各单体间的界面上。

这个锌离子对于激活水分子是必需的，它是通过将其 pK_a 降低至 7 左右来完成对水分子的激活。 于是在活化形式的碳酸酐酶中存在着一个结合在锌离子上的羟基离子。 反应生成的质子可以与催化反应相同的速率被释放到外部的水或缓冲液中。 为了做到这一点，活性最高的碳酸酐酶活性位点中有一个组氨酸残基作为临时的缓冲分子。 这个组氨酸残基并没有以氢键形式结合任何其他的残基并可以自由地旋转以释放反应生成的质子。 该组氨酸残基并不直接处于与锌离子结合的水分子旁边，而是以一些水分子作为桥来传递质子。

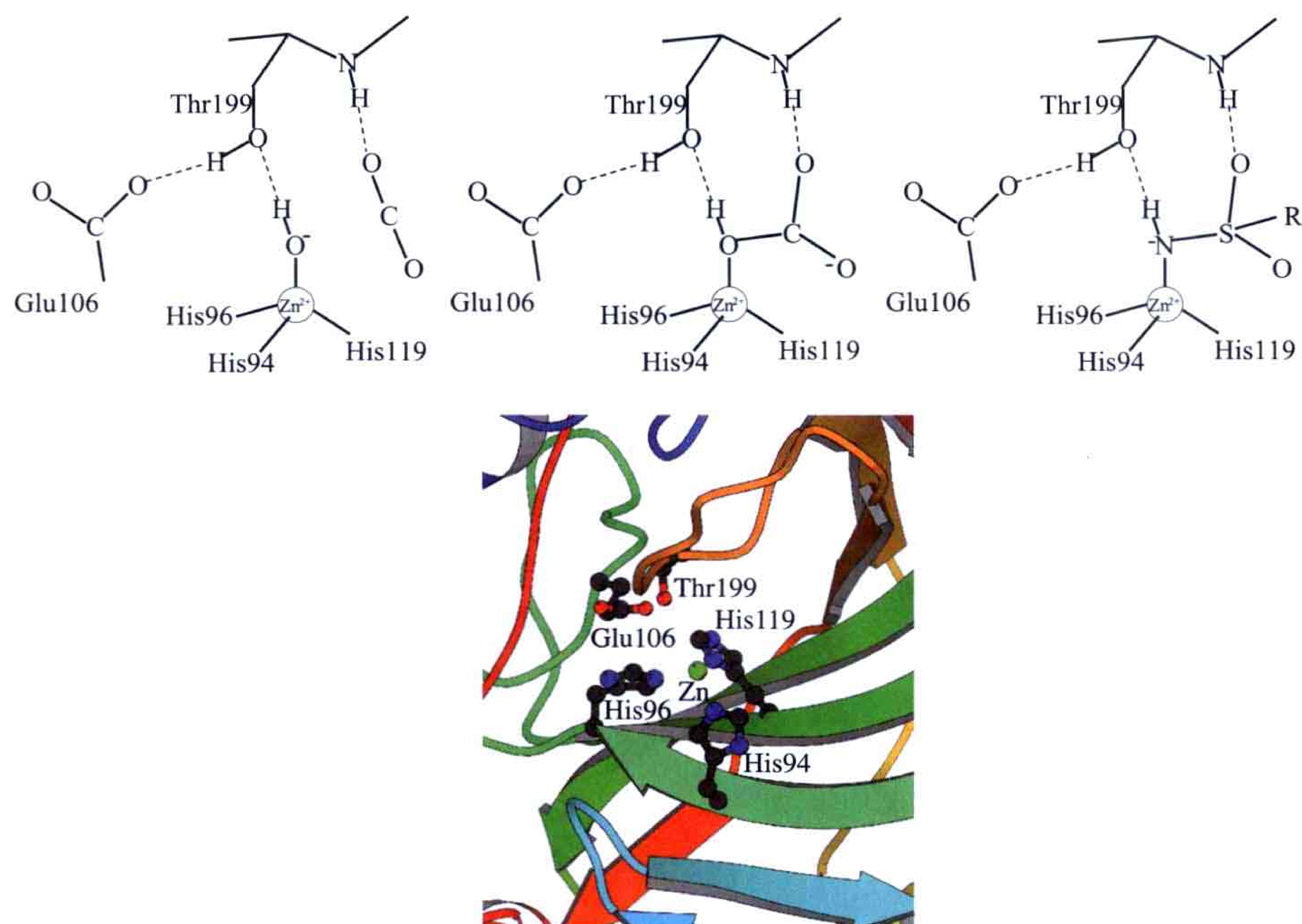


图 5.2 α 型碳酸酐酶的活性位点。上左：羟基离子已经处于准备进行亲核攻击的状态。上中：碳酸氢根离子的质子化的氧而非带负电荷的氧结合在碳酸酐酶的锌离子上。这是因为碳酸酐酶 199 位苏氨酸的羟基与 106 位谷氨酸形成了氢键。于是 199 位苏氨酸羟基的位置及朝向阻止了非质子化的基团与锌离子结合。这个催化反应的速率仅仅取决于羟基和碳酸氢根离子之间化学键的形成或打断。上右：碳酸酐酶与磺胺类抑制剂（sulfonamide inhibitor）结合的模型。抑制剂的 NH 基团上质子化的氢结合在碳酸酐酶 199 位的苏氨酸上，从而使整个带负电荷的 NH 基团被摆放在最适合的锌离子第四配位的位置。下：未结合底物的碳酸酐酶活性位点示意图（PDB：2CBB）。

碳酸酐酶活性中心口袋的大小对底物及产物的扩散速率没有限制作用。为了使催化反应快速进行，酶既不应该强力地结合底物或者产物，也不应该进行大的构象变化（conformational change）。这正是碳酸酐酶反应如此快速的秘密。碳酸酐酶迫使底物以一种出乎意料的方式与其结合。在锌离子附近存在着一个替补的氢受体，使得只有质子化的配体才可以结合在锌离子上（图 5.2）。这个被称为“守门员”的氢受体便是 199 位苏氨酸的羟基，这个羟基中的氢与 106 位谷氨酸形成氢键而被牢牢地占据着。这就防止了碳酸酐酶的底物，即碳酸氢根以带负电荷的氧与锌离子结合。取而代之的是碳酸氢根的 OH 基团与锌离子结合。这个碳酸氢根 OH 基团与碳原子之间的化学键被破坏，或者其形成不需要任何底物或产物的重新排列，也不需要底物质子的转移。此外，碳酸氢根转换成二氧化碳的过程也不需要碳酸酐酶进行任何构象变化。

在生成碳酸氢根时，羟基离子对二氧化碳分子进行亲核攻击。负电荷被转移到二氧化碳分子中离锌离子最近那个氧上，而这个氧原子可能经过了一个微小的取向变化后变成锌离子的一个远端配体。碳酸氢根通过这种中性的、质子化的氧结合在锌离子上，结合力不是很强，于是它会很快被一个水分子所取代。水分子结合后被质子化成羟基离子，从而导致了产物即碳酸氢盐的释放。

5.1.1　中间状态的稳定

芳香磺酰胺（aromatic sulfonamide）是所有碳酸酐酶的一种强效抑制剂（图 5.2）。这种抑制剂属于酶催化反应的中间产物类似物。总的说来，酶是通过在反应路径中生成中间产物降低化学反应能量壁垒的方式来实现催化的。化学反应的中间状态（transition state）是一种高能状态，反应物必须穿越这个状态以生成产物。许多酶都通过稳定这种中间状态来降低反应的能量壁垒从而促进反应进行。在碳酸酐酶催化的反应中，与锌离子结合的带负电荷的羟基离子和二氧化碳发生反应以生成碳酸氢根。芳香磺酰胺的磺胺基团则是其中间产物的类似物。带负电荷且质子化的 NH 基团与锌离子结合并且作为氢供体与 199 位苏氨酸的 OH 形成氢键，同时磺胺基团的两个氧原子以一种类似于二氧化碳或者是碳酸氢根中的两个非质子化氧的方式结合在锌离子上。另外，磺酸并不能作为碳酸酐酶的抑制剂，因为磺酸根（SO_3^-）中缺少质子。

碳酸酐酶的催化反应速率如此之快的原因除了这个反应本身十分简单以外，碳酸酐酶中锌离子、199 位苏氨酸及 106 位谷氨酸的排列方式也功不可没。如果将 199 位的苏氨酸点突变成其他氨基酸，可以发现碳酸氢根与锌离子的结合会更加紧密，并且其两个氧原子成为了锌离子的配体。这就在很大程度上降低了碳酸酐酶的催化效率。同样，如果突变了 106 位的谷氨酸，碳酸酐酶的催化效率也会大大降低。

尽管 α 型、β 型和 γ 型的碳酸酐酶有着非常不同的结构，并且在活性位点上的残基也不尽相同，但锌离子始终是所有碳酸酐酶的主角。在 β 型碳酸酐酶中与锌离子结合的配体与另两类碳酸酐酶有些许不同，但磺胺类抑制剂同时是所有三种类型碳酸酐酶的强效抑制剂，这意味着这三种类型的碳酸酐酶有着相同的中间状态以及催化机制。因此可以说碳酸酐酶的存在是趋同演化（convergent evolution）的一个例子——就像车轮在历史上曾经被独立地发明了三次以上一样。

5.2　核苷酸还原酶—— 一种被高度调控的酶

核苷酸还原酶（ribonucleotide reductase, RNR）催化的反应是将核苷酸的第 2 位的羟基替换成氢原子从而生成脱氧核糖核苷酸（deoxyribonucleotide）（图 5.3）。这正是合成 DNA 所需的原料被从头制作出来的过程。这个酶的演化是 RNA 世界转化为 DNA 世界的一个关键性步骤。生成不同的 dNTP 或 dNDP 是直接通过 RNR 的变构调节机制来完成的。蛋白质中的一个被转变为含硫自由基（thiyl radical）的半胱氨酸残基是催化反应的中心组件。

核苷酸还原酶有三种类型，即 Ⅰ 型、Ⅱ 型和 Ⅲ 型（图 5.4）。它们的区别在于与氧原子的相互作用方式及生成稳定或临时的自由基以便生成含硫自由基的方式（表 5.1）。对于多亚基的 Ⅰ 型及 Ⅲ 型核苷酸还原酶来说，含硫自由基是由一个不参与催化的活化

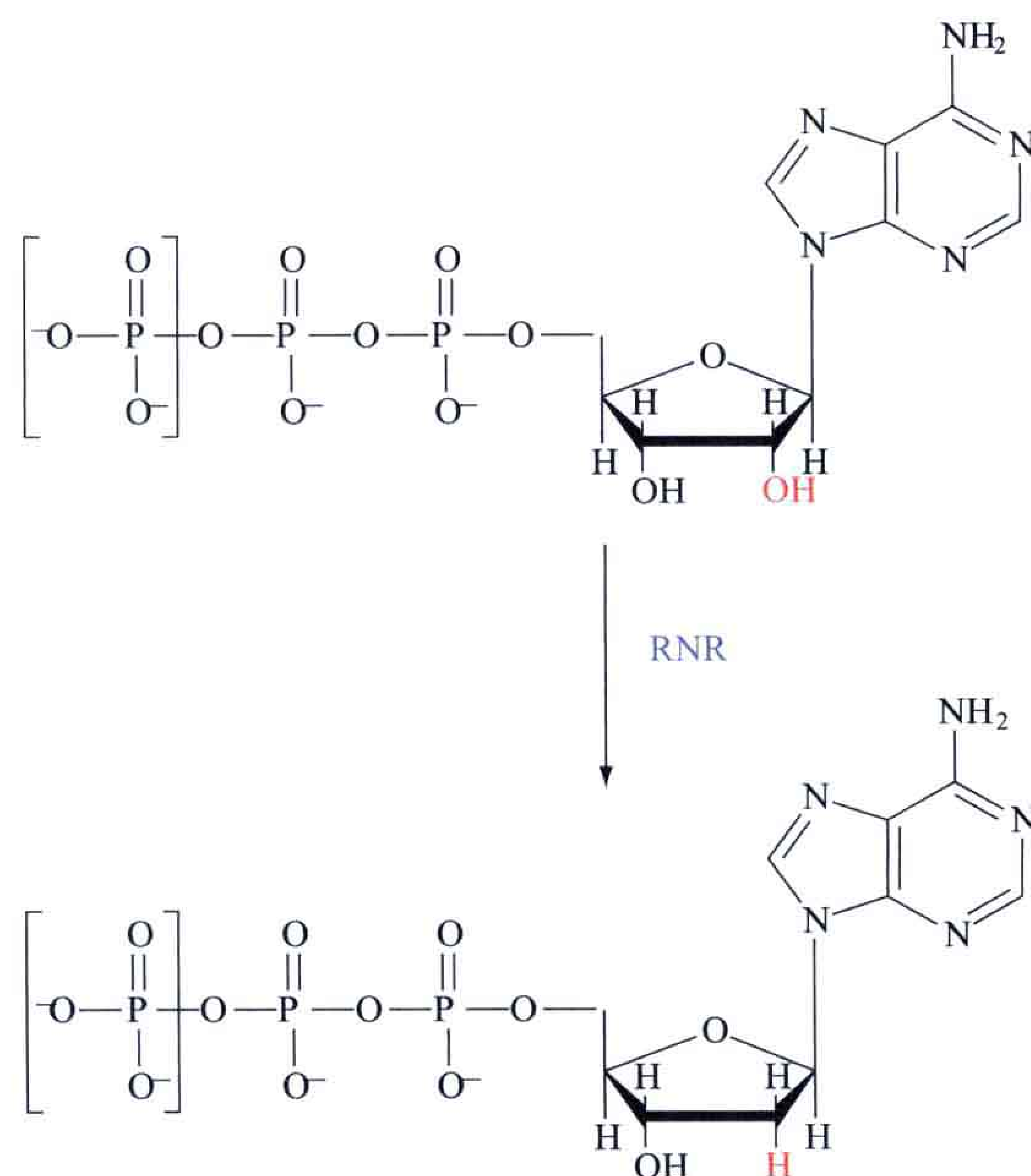


图 5.3　由核苷酸还原酶所催化的反应。ATP 的核糖环上 2′ 位的羟基基团被替换为氢原子从而形成了 dATP（特别感谢 Derek Logan 提供此图）。

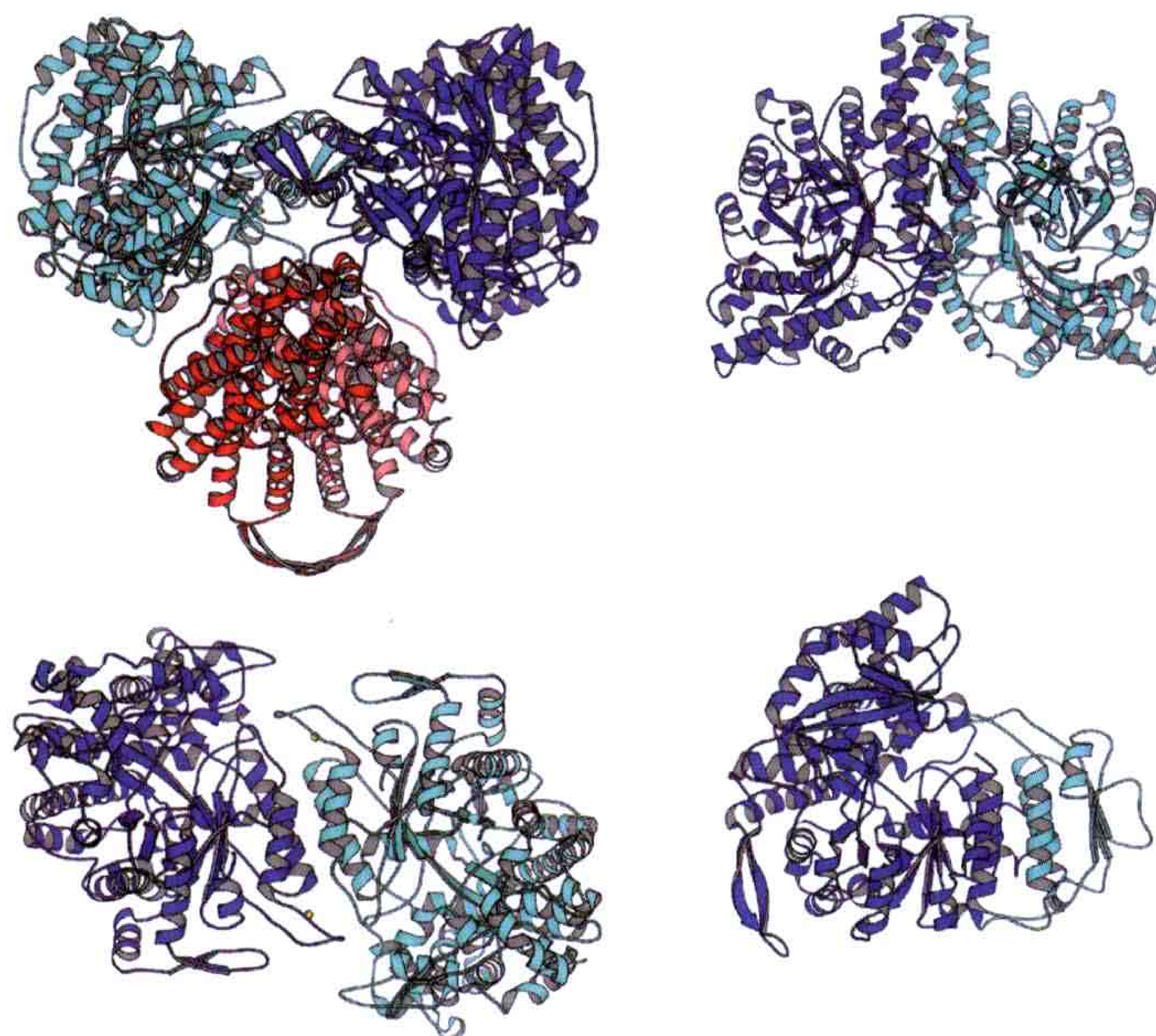


图 5.4　三种不同类型的核苷酸还原酶。上左：来自于大肠杆菌（*E. coli*）的四聚体形式的 I 型核苷酸还原酶。这是一种之前从未发现过类似对称性的复合体（基于 PDB：1RLR 和 1RIB）。上右：T4 噬菌体（phage T4）的Ⅲ型核苷酸还原酶是四聚体形式，这是其中 α 催化亚基二聚体结构（PDB：1HK8）。Ⅲ型核苷酸还原酶 β 亚基的结构目前尚未解出。下左：来自海栖热袍菌（*Thermatoga maritima*）二聚体形式的Ⅱ型核苷酸还原酶（PDB：1XJE）。下右：来自赖氏乳酸杆菌（*Lactobacillus leichmannii*）单体形式的Ⅱ型核苷酸还原酶（PDB：1L1L）。催化亚基均为蓝色和青绿色，而 I 型核苷酸还原酶的 R2 亚基为红色。

亚基生成的；而Ⅱ型的核苷酸还原酶生成含硫自由基的方法是直接在催化亚基上切割腺苷酰钴胺素（adenosylcobalamin）得来。Ⅰ型核苷酸还原酶生成自由基需要氧的帮助而Ⅱ型不需要氧；对于Ⅲ型核苷酸还原酶来说，生成甘氨酸自由基反而会被氧所抑制（表 5.1）。因此三种不同类型的核苷酸还原酶的分化可能取决于在演化过程中氧气的存在与否。

表 5.1　不同类型核苷酸还原酶的特性

特性	Ⅰ型	Ⅱ型	Ⅲ型
寡聚化状态	$R1_2 * R2_2$	α 或 α_2	$\alpha_2\beta_2$
需氧情况	需氧	不需要氧	厌氧（氧造成损害）
首要自由基	酪氨酸	腺苷酰钴胺素	甘氨酸
自由基生成处	铁 - 氧 - 铁中心	腺苷酰钴胺素	[4Fe-4S] 中心
			硫代腺苷
			甲硫氨酸，还原态的黄素氧化还原蛋白
电子供体	硫还原蛋白或谷氧化还原蛋白	硫还原蛋白或谷氧化还原蛋白	甲酸盐

核苷酸还原酶可以还原所有的 4 种 NTP 或 NDP 生成相应的 dNTP 或 dNDP。而仅靠核苷酸还原酶一个酶就可以维持生物组织内 4 种脱氧核糖核苷酸的平衡供给。 酶活性的调节原则也十分简单：哪种脱氧核糖核苷酸短缺，就合成哪种脱氧核糖核苷酸。这种调节是通过一个调控位点来完成的，这个位点可以结合 NTP 或 dNTP，并由此调节酶的底物特异性（表 5.2）。 除此之外，一些类型的核苷酸还原酶还含有一个普适的激活位点，这个位点可以结合 ATP 或 dATP。结合 ATP 可以激活酶；相反，结合 dATP 可以抑制酶活性，这样可以减少核苷酸还原酶对 dNTP 的依赖性。这种活性调控的结构背景正在逐渐浮现出来。

表 5.2　核苷酸还原酶的变构调节

调节位点	最适底物		
	Ⅰ型	Ⅱ型	Ⅲ型
dTTP	GDP	GDP	GTP
dGTP	ADP	ADP	ATP
dATP	CDP/UDP	CDP/UDP	CTP
ATP	CDP/UDP	CDP/UDP	CTP

5.2.1　核苷酸还原酶的结构

三种不同类型的核苷酸还原酶有着相似的三维结构，并且参与变构调节的二级结构也十分类似。 因此这三种不同类型的核苷酸还原酶必然是由同一个祖先经过趋异演化（divergent evolution）而来的（图 5.5）。所有类型的核苷酸还原酶催化亚基的核心都是一个 10 股 α/β 桶结构（10-stranded α/β barrel）。这个桶是由两个 5 股的平行 β 片层以反平行的方式组成的。

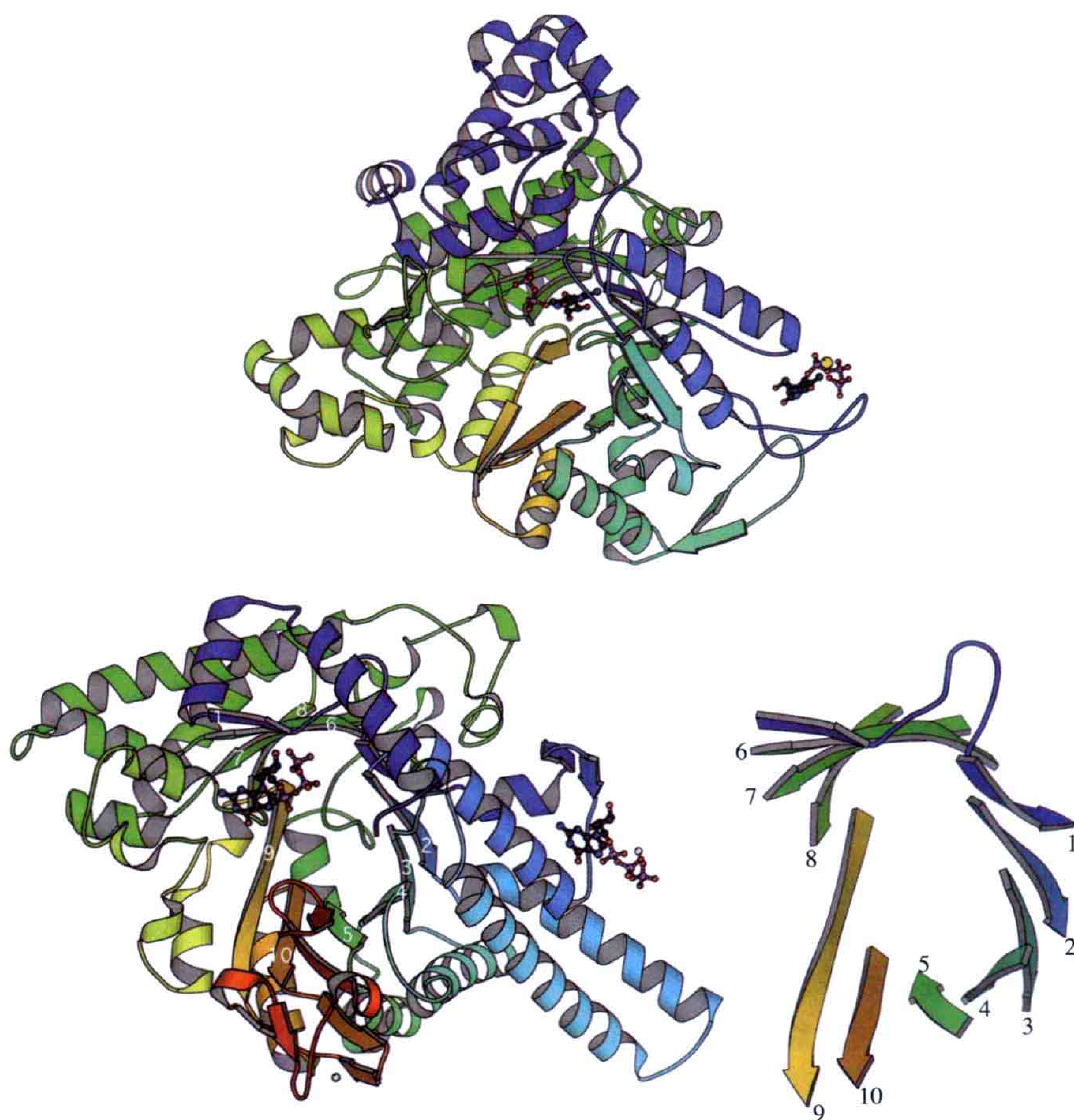


图 5.5 核苷酸还原酶催化亚基的结构。上方所示为来自海栖热袍菌的Ⅱ型核苷酸还原酶（PDB：1XJE），这个结构代表了Ⅰ型和Ⅱ型的核苷酸还原酶。而下方所示为来自 T4 噬菌体的Ⅲ型核苷酸还原酶（PDB：1HK8）。从图中的结构可以明显看出三种类型的核苷酸还原酶在演化上是有关联的。桶状结构的每一股的编号如下右方所示，在此只显示了桶状结构中的 β 股拓扑结构。 核苷酸还原酶的活性位点在图中以结合核苷酸的形式显示，即处于 β 桶状结构的中心，而调控位点位于图中结构的右侧。

5.2.1.1 生成自由基的位点

三种类型的核苷酸还原酶产生含硫自由基的方式是不同的，并且是由使用氧的方式决定的。 在Ⅰ型核苷酸还原酶中，自由基最初是由一个靠近铁 - 氧 - 铁复合体（Fe-O-Fe complex）的酪氨酸残基生成的，而这个铁 - 氧 - 铁复合体被包埋在酶的 R2 亚基中。这个自由基必须从 R2 亚基转移（或许作为电子 - 质子的协同转移）到活性中心去，而活性中心所在的 R1 亚基离 R2 亚基有 30Å 的距离。 这个转移的机制一直被广泛讨论却始终没有得到一个肯定的结论。 其他类型的核苷酸还原酶不需要进行如此大距离的自由基转移。 在Ⅱ型核苷酸还原酶中，钴胺素（cobalamin）很可能直接与半胱氨酸残基作用从而生成含硫自由基。而在Ⅲ型核苷酸还原酶中，甘氨酸自由基也离活性位点很近。

5.2.1.2　酶活性的调节

如前所述，核苷酸还原酶含有一个（有时两个）调节酶活性的位点，以保证细胞内4种脱氧核糖核苷酸的平衡供给。（不受调节的）总体活性位点（overall active site）不是一直都存在着（图5.6）。当ATP结合该位点时酶的活性提高，而当dATP结合该位点时酶的活性被抑制。这种调节机理仍然不是十分清楚。

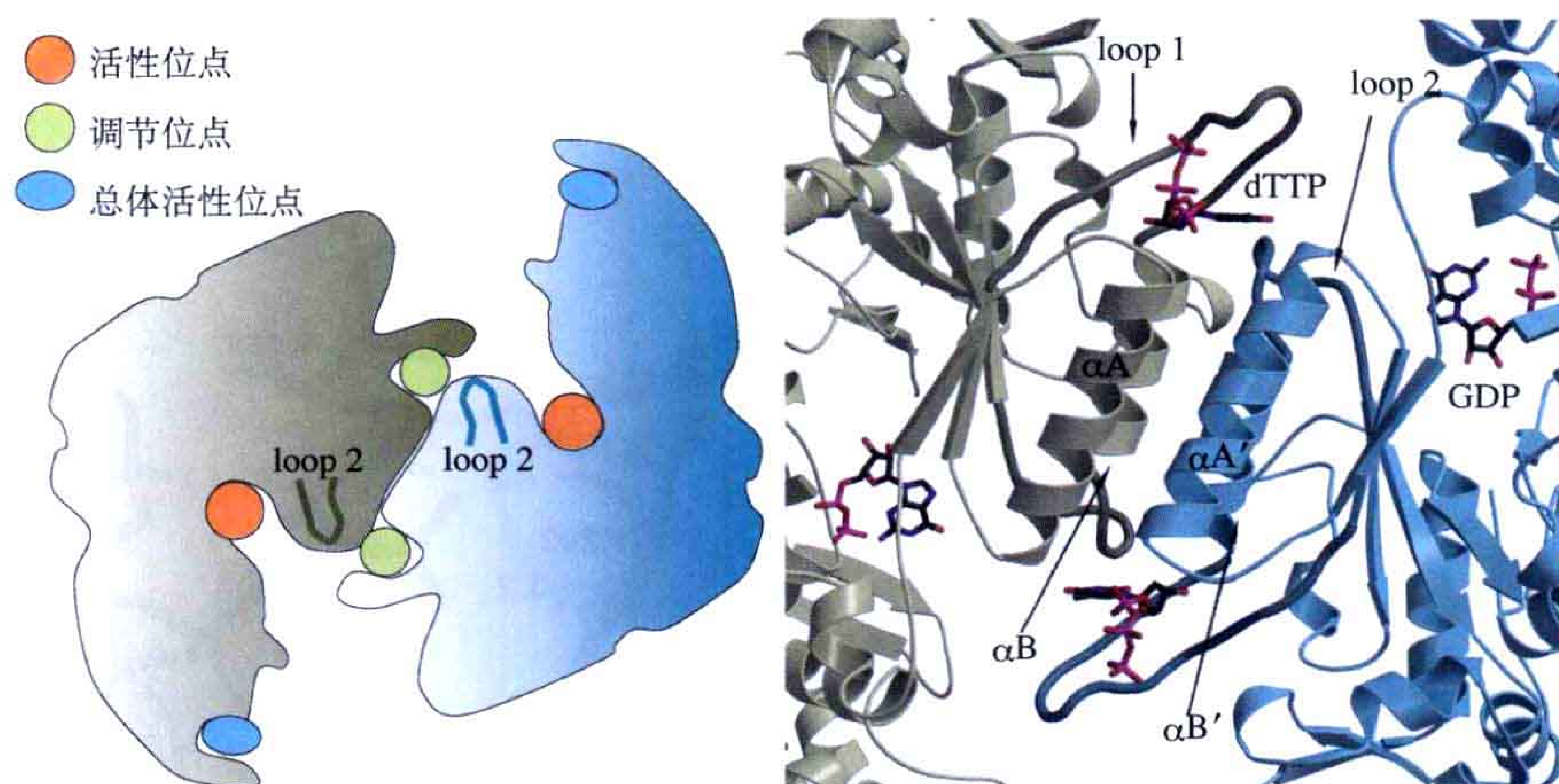


图5.6　左：二聚体形式的I型核苷酸还原酶催化亚基示意图。图中可以看出活性位点与调节位点的空间位置关系。不受调节的总体活性位点（overall activity site）只存在于一些I型和III型核苷酸还原酶中。右：效应物分子（dTTP）与底物分子（GDP）结合位点的结构细节。在此感谢Derek Logan提供此图。值得注意的是loop2位于调节位点和催化活性位点的两个核苷酸中间，而两个核苷酸的碱基均朝向loop2。loop2在调节底物的选择性方面扮演着很重要的角色。以上相关位点的排列方式在其他类型的核苷酸还原酶中也是类似的。

在I型核苷酸还原酶的一个亚基内，这个调控底物特异性位点或效应物位点是远离活性位点的。然而，在二聚体形式中这个调控位点跨过亚基的界面从而处于活性位点的对面(图5.6)。乍看之下这样的安排似乎无法适用于单体形式的II型核苷酸还原酶，但是在这些酶中，存在一个小的插入结构域（inserted domain）模拟了所缺失的亚基上的调节位点的一些必要部件（图5.4）。最确凿的实验证据来自于核苷酸同时结合在两个位点（调节位点和活性位点）上的实验。这个实验的研究对象是二聚体形式的II型核苷酸还原酶，而研究的小分子对分别是dTTP-GDP、TTP-GDP、dGTP-ADP、dATP-CDP、dATP-UDP、AMPPNP-CDP及AMPPNP-UDP。

在不同的复合物中，两个核苷酸的糖环和磷酸均以相同的方式与蛋白质结合，并且与蛋白质的相互作用也一样。这样就提高了处于效应位点的核苷酸影响核苷酸结合到活性位点的可能性。loop1仅与结合在效应位点的核苷酸相互作用。而处于αB和βC之间的环（即loop2）则是影响效应位点的核苷酸与底物结合位点关系的关键性元素。loop2（以海栖热袍菌的II型核苷酸还原酶为例，即第199位至210位的残基）的位置横越两个亚基的界面并且处于效应位点和活性位点之间。这个环具有很强的柔性，当

结合在效应位点时，不同的核苷酸与不同的主链基团结合从而导致较大的构象变化（图5.7）。这种构象变化导致了活性位点对于底物的不同倾向性（图 5.8）。在所有的复合体中，底物是被第 210 位的丙氨酸所固定的。第 207 位精氨酸的胍基有些时候（但并不总是）与底物的碱基形成堆积作用，并且与磷酸形成电荷相互作用。 202 位的赖氨酸和 203 位的谷氨酰胺是两个关键性的残基，它们与底物的碱基形成氢键，并且对于不同的底物有着不同的形成氢键的方式。 对于 loop2 的构象比较及这些关键残基的位置可见图 5.9。 而对于这种调节的机理，至今仍然没有一个完整而合理的说明。

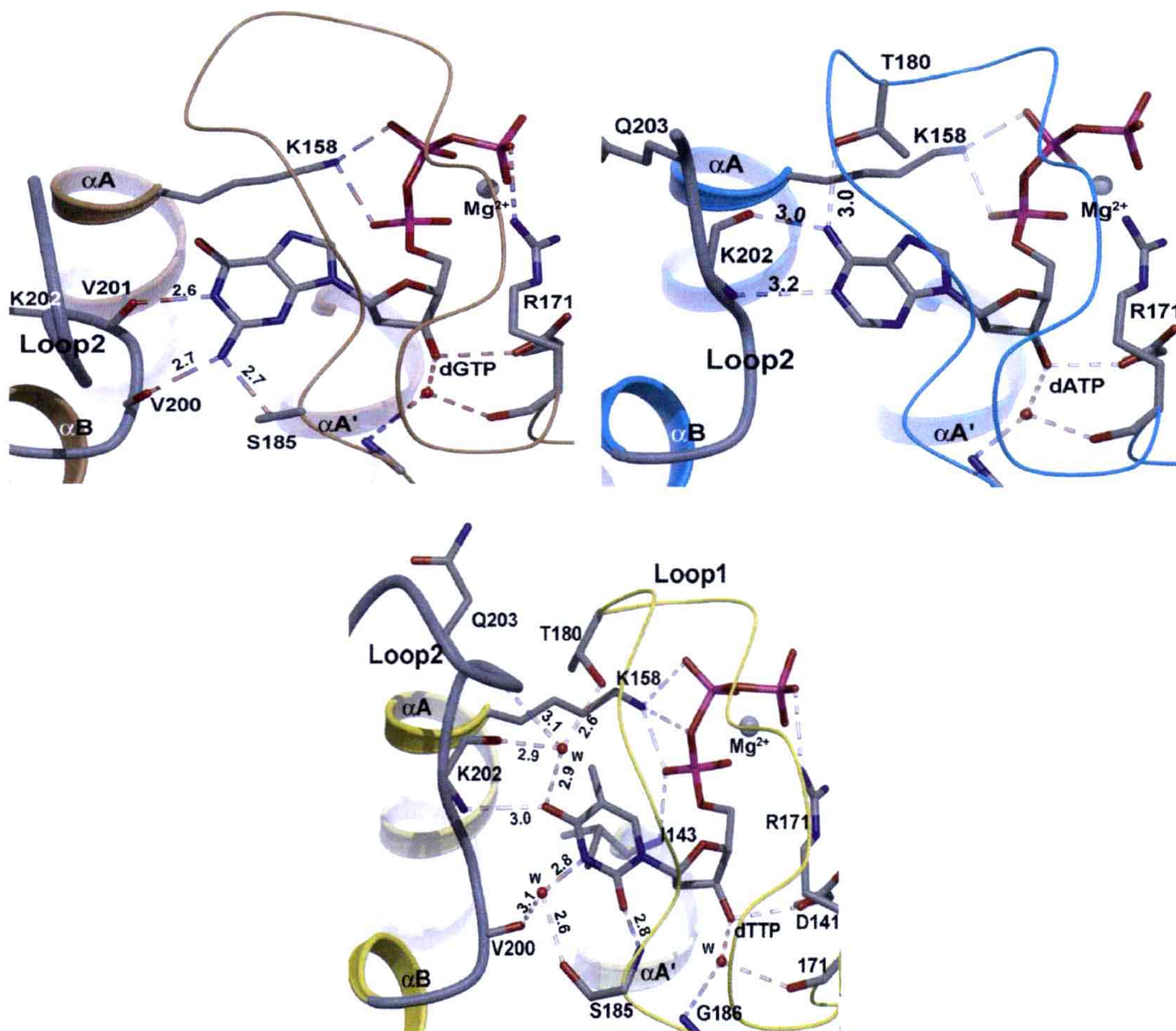


图 5.7 dNTP 结合在Ⅱ型核苷酸还原酶的效应位点（dGTP、dATP 和 dTTP）。脱氧核糖以及磷酸均以一种相似的方式结合在效应位点上，并将核苷酸固定在一个特定的位置。 这样的结果是核苷酸的碱基以非常不同的方式与 loop1 和 loop2 相互作用，而 loop1 和 loop2 所采取的构象也相应地有着显著差异。效应分子通过这种方式将其结构信息传递给活性位点，从而决定哪种核苷酸可以与活性位点结合并发生还原反应 [PDB：1XJJ、1XJF 和 1XJM，在此感谢 larsson KM 等提供刊印的授权，图片原文出自 (2004) Structural mechanism of allosteric substrate specificity regulation in a ribonuleotide reductase. *Nature Struct Mol Biol* **11**:1142-1149. 版权所有：(2004) *Nature*]。

图 5.8 Ⅱ型核苷酸还原酶的活性位点对于底物的选择。正如预期，核糖和磷酸以完全相同的方式结合在活性位点上。对底物的选择性则是由 loop2（199 位至 210 位的残基）所决定的。而 loop2 的这种选择性是由结合在效应位点（或者说是底物选择性位点）上的核苷酸决定的。有趣的是，核苷酸还原酶对 GDP 的识别完全是通过主链原子完成的。在这个复合体中，loop2 的底物一侧是完全有序的。在另一个复合体中，另一些来自于 loop2 效应位点一侧的残基指向了活性位点，而 loop2 靠近活性位点的一侧则是无序的（PDB、1XJE、1XJN、1XJK 及 1XJG）[在此感谢 larsson KM 等提供刊印的授权，图片原文出自 (2004) Structural mechanism of allosteric substrate specificity regulation in a ribonuleotide reductase. *Nature Struct Mol Biol* **11**:1142-1149. Copyright (2004) *Nature*.]。

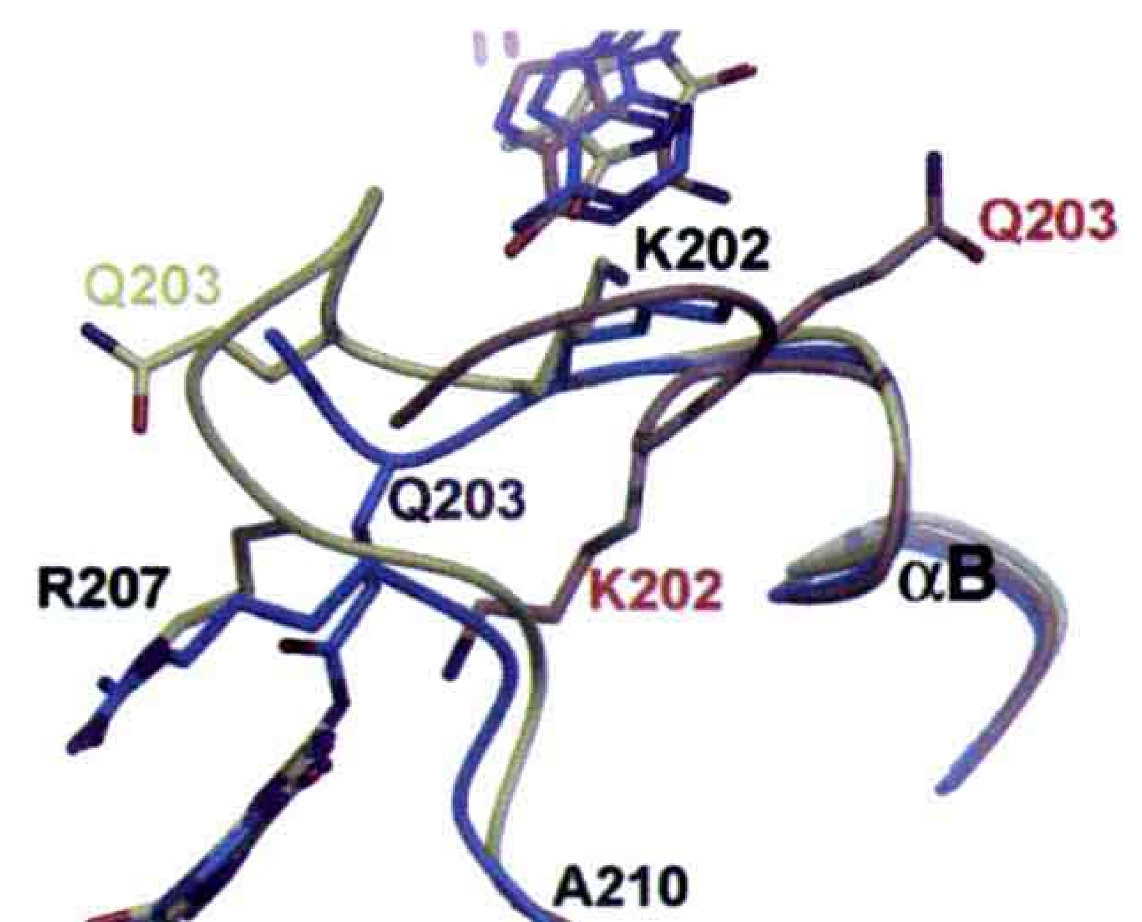


图 5.9 分别结合三种不同的效应分子时 loop2 与处于效应位点和活性位点之间的关键残基（即 202 位赖氨酸及 203 位谷氨酸）的构象比较（PDB：1XJN 和 1JXK，不同颜色所代表的复合体与图 5.8 一致）[在此感谢 larsson KM 等提供刊印的授权，原文出自 (2004) Structural mechanism of allosteric substrate specificity regulation in a ribonuleotide reductase. *Nature Struct Mol Biol* **11**:1142-1149. Copyright (2004) *Nature*.]。

5.2.1.3 活性位点与催化

对于Ⅱ型核苷酸还原酶来说，4种不同底物的磷酸和核糖部分与活性位点的结合均采用几乎完全相同的方式。C439位半胱氨酸所携带的含硫自由基正处于核糖的3′位碳原子附近(图5.10)。在核糖的另一边，有两个半胱氨酸（C225位和C462位）相互靠近，其距离足以形成二硫键。核糖上的两个羟基基团与N437位的天冬酰胺形成氢键，而3′位的羟基又与E441位的谷氨酸形成氢键。含硫自由基通过吸收核糖3′碳上的氢将自由基转移到核糖上。这导致了2′位的羟基变成水分子离去，而变成水分子所需的那个质子由两个靠近的半胱氨酸中的一个提供。接下来，核糖又会从两个半胱氨酸处重新获得一个氢原子，而此时这两个半胱氨酸形成二硫键，且重新获得了自由基。二硫键需要被还原以便能继续参加下一循环的反应。

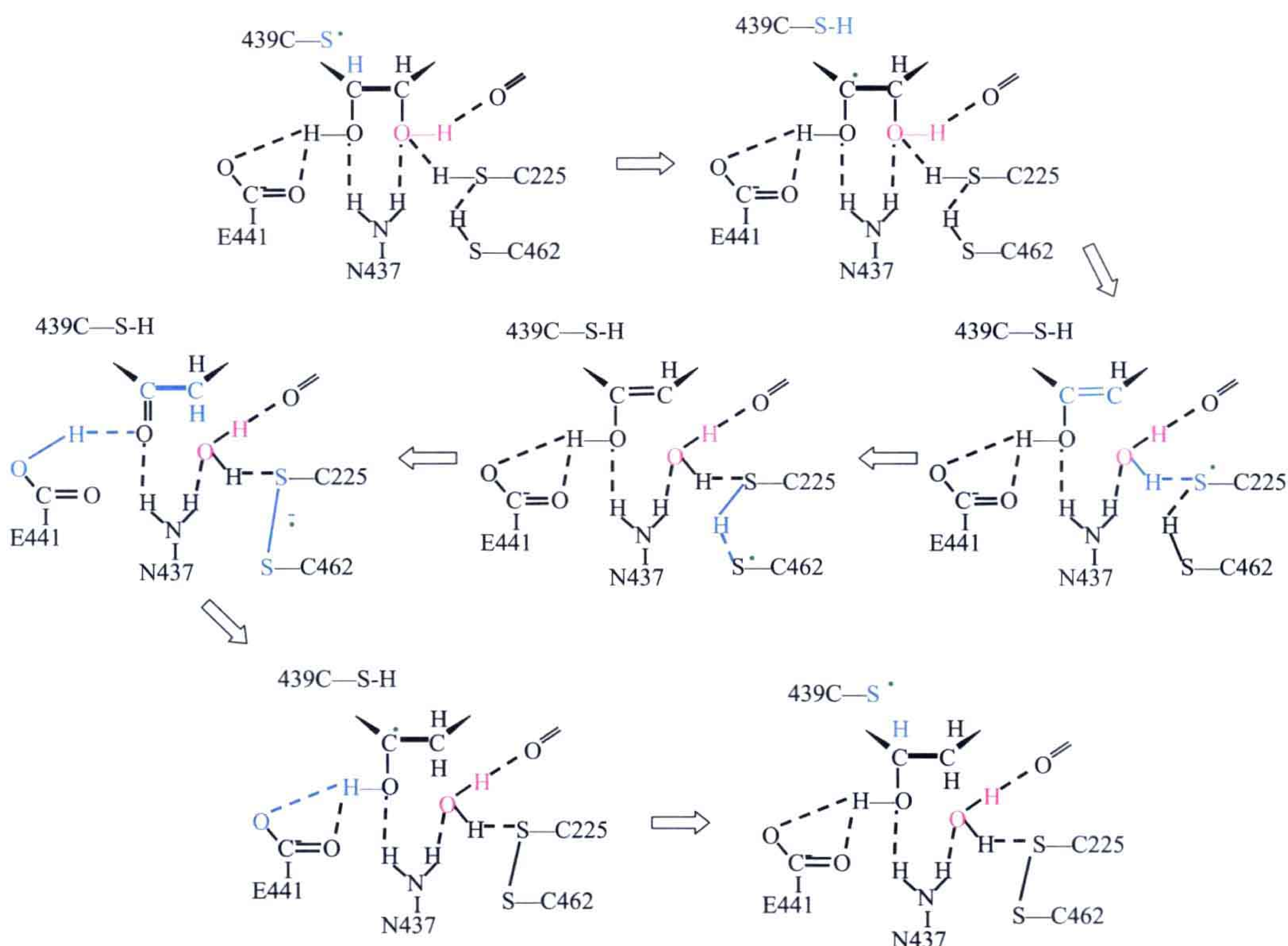


图5.10 核苷酸还原酶的催化机理（2006年由Nordlund和Reichard提出）。自由基显示为绿色小点，核糖上的羟基显示为红色，与下一个步骤相比发生变化的原子和化学键显示为蓝色。图中所有关于残基的编号均是以Ⅰ型核苷酸还原酶为模板，故本图中第437、439及441位残基对应于图5.8即Ⅱ型核苷酸还原酶的第320、322和324位残基。

5.3 马达蛋白和分子开关

ATP是细胞的“能量货币”。在生物组织中需要大量的ATP消耗。一个成年人

一天要消耗 50~75kg 的 ATP。ATP 是通过跨膜的化学梯度差产生的。 这种梯度差的产生一般是源自于膜两边不同的 pH 或钠离子浓度（详见第十章）。这种梯度差可以被用来产生运动，即动能，而马达蛋白就可以利用这种能量产生 ATP 形式的化学能。反之，ATP 水解的化学能又可被用来驱动一系列的分子过程。 这种化学能可以被用来产生动能。ATP 水解酶（ATPase）负责水解 ATP，并且在水解 ATP 后会生产构象变化，而被用来完成不同种类的工作。

GTP 水解酶（GTPase）是另一种水解核苷三磷酸的酶，它水解 GTP 以完成相应的工作。GTP 水解酶，又称 G 蛋白（G-protein），被广泛当做分子开关使用。 G 蛋白通常有两种主要的状态，称为“开”（ON）和“关”（OFF），用来告知整个系统它们的任务是已经完成了还是正等待完成。 也许有人会问是否有一些 GTP 水解酶也可以行使马达蛋白的功能。 区分马达蛋白和分子开关的一个重要特点在于核苷酸是何时被水解的。 经典的马达蛋白会在构象变化前，即在工作完成前水解 ATP 分子。而分子开关在“开”的状态时会诱导一个过程，当这个过程完成的时候，核苷酸才被水解，从而导致分子开关的构象变化，于是造成了分子开关与其受体的亲合性下降。ATP 水解酶和 G 蛋白所催化的步骤均是不可逆的，这是由三磷酸水解的特性决定的。

5.3.1　具有 P-loop 的核苷三磷酸水解酶类

马达蛋白和 G 蛋白同属于一个大的蛋白质超家族（superfamily），这个家族的特点是通常使用一种环（P-loop）来结合核苷酸。 P-loop 的一级序列十分保守，为 GXXXXGKT/S。这些蛋白质都含有一个位于中心的、几乎为平行排列的 β 片层，而每个 β 股之间由 α 螺旋来连接。通过独特的 β 片层拓扑结构及一些其他元素可以区分出不同的家族成员（表 5.3）。

表 5.3　一些含有 P-loop 的蛋白质或者蛋白质结构域

蛋白质家族及其 β 片层的拓扑结构 [a]	蛋白质或蛋白质结构域	章节 [b]
核苷及核苷酸激酶：β 股顺序为 23145	胸腺嘧啶核苷激酶	16.3
	脱氧核糖核苷酸激酶	16.3
	脱氧胞苷激酶	16.3
马达蛋白：β 股顺序并非完全保守	肌球蛋白的重链	12.2.3
	驱动蛋白的重链	12.3.2
固氮铁蛋白类：β 股顺序为 3241567	信号识别蛋白 FfH，GTP 水解酶结构域	—
RecA 类：β 股顺序为 32451678，β 股 7 是反平行的	细菌 recA 蛋白，ATP 水解酶结构域	6.1.5
	F1 ATP 合酶的 α 和 β 亚基（中心结构域）	5.3
	DNA 和 RNA 解旋酶	6.1.1
	转录终止子 Rho	6.1.1
ABC 转运蛋白；含有额外亚结构域的 RecA 类蛋白	参与“主动运输”的 ATP 结合蛋白（ABC 转运蛋白）	10
串联 AAA-ATPase 结构域：两个重复的 RecA 类结构域	DNA 解旋酶	—
	RNA 解旋酶	—
	RecQ 解旋酶结构域	6.1.1

续表

蛋白质家族及其β片层的拓扑结构[a]	蛋白质或蛋白质结构域	章节[b]
延伸的 AAA-ATPase 结构域（AAA+）：β股顺序为 23415，C 端形成螺旋束	核苷酸修复蛋白 UvrB	6.1.1
	eIF4a	—
	DNA 聚合酶Ⅲ，γ 亚基	6.1.2
	Holliday 交叉解旋酶，RuvB	—
	HsIU，Lon，FtsH	9.2.3
	Lon	
	ClpA，ClpB	9.1.2.5
病毒 RNA 解旋酶	登革热病毒解旋酶	—
Ⅱ型胸腺嘧啶核苷激酶：β股顺序为 324516	胸腺嘧啶核苷激酶 2	16.3
G 蛋白：β股顺序为 231456，β股 2 是反平行的	单体 G 蛋白（Ras）	5.3，11.3.2
	三聚体 G 蛋白，α 亚基（Gi，Gs，Gt）	11.4.2
	参与翻译的 G 蛋白（EF-Tu，EF-G 等）	8.5

a 分类依据 SCOP 数据库（详见附录 B.2.2.1）。
b 章节是指此蛋白质在本书中被讨论到的章节。

最常见的 NTP 水解酶是一种被称为“Walker 型”（名字来源于最早提出此序列模体的 John Walker，他毕生从事 ATP 水解酶及合酶相关的研究并因此而获得 1997 年诺贝尔化学奖，见附录 F——译者注）的 NTP 水解酶。它们拥有一些序列比较保守的模体，分别被称为 Walker A（即 P-loop）和 Walker B（4 个疏水的残基后面紧跟着一或两个酸性残基）。这些酶分为一些不同的类型，我们将会在下面对其中两个进行讨论。P-loop 均位于一段 β 股和一段 α 螺旋的连接处，且与 ATP 或者 GTP 的 α 位及 β 位的磷酸相互作用。尤其是 P-loop 上的一些残基的主链氮原子会与 β 磷酸相互作用（详见 5.3.3.1 节的图 5.23 和图 5.24）。

ATP 水解酶中有一个很大的家族被称为 AAA+ 蛋白，即与各种细胞活性相关的 ATP 水解酶（ATPase associated with various cellular activity）。真核生物体内至少含有 50~80 种 AAA+ 蛋白。它们都是寡聚化形式的 ATP 水解酶，并且参与各种不同类型的工作。AAA+ 蛋白经常形成六聚的环状结构并参与组成大型的分子聚合物。它们可以参与组成分子伴侣（详见第九章），帮助没有正确折叠的蛋白质进行重新折叠；或者参与组成需 ATP 的蛋白质水解酶，如蛋白酶体来降解错误折叠或者是不需要的蛋白质（详见第九章）。它们也可以作为解旋酶来解开核酸的双螺旋或者参与将病毒基因组打包进病毒外壳的分子机器。ATP 合酶（ATP synthase）则是 AAA+ 蛋白类型中最核心的一种酶。

5.3.2 ATP 合酶

ATP 是在线粒体、叶绿体及细菌中合成的。ATP 合酶是一种独一无二的蛋白质，它在从细菌到人类的生物体中都是保守的。而由 ATP 合酶所催化的将 ADP 和磷酸合成为 ATP 的反应，是所有细胞中发生得最多的生理学反应。如图 5.11 所示的便是这个参与产生离子梯度及合成 ATP 的系统。

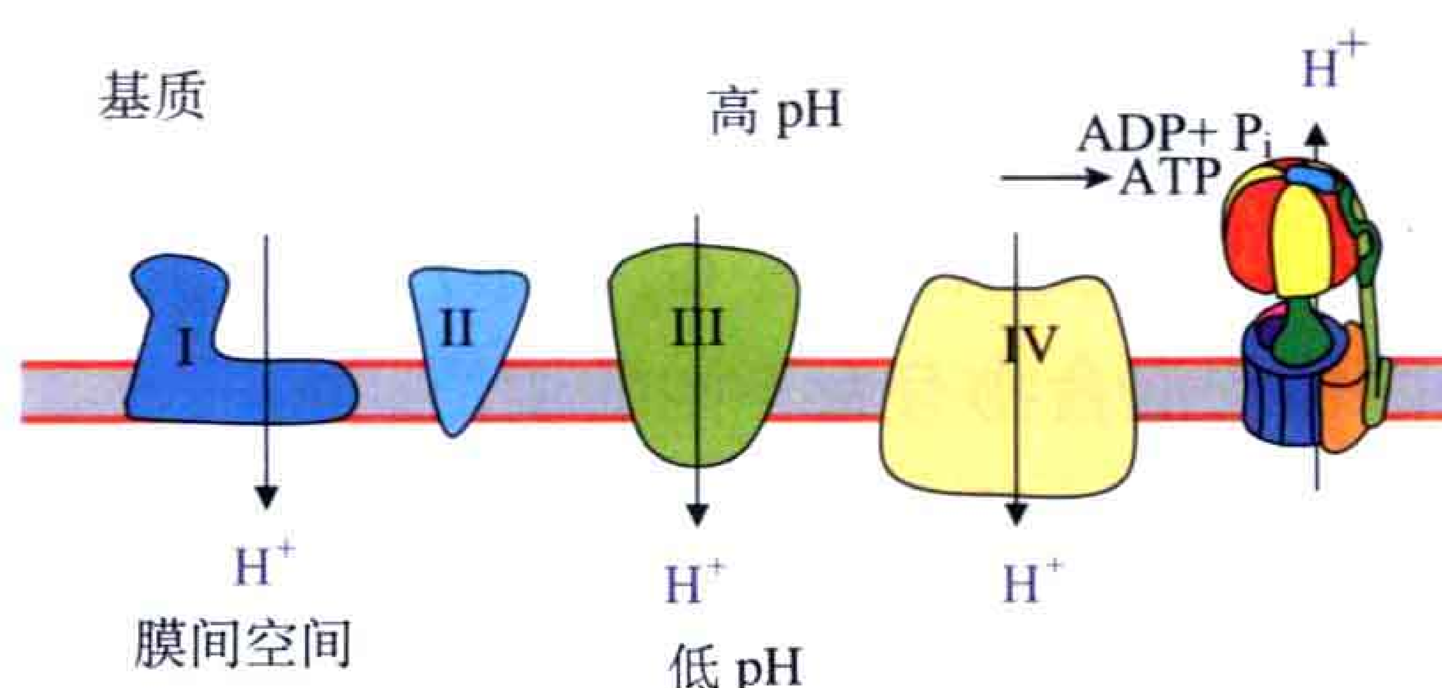


图5.11　在呼吸链中，质子被复合体 I（NADH脱氢酶）、复合体Ⅲ（细胞色素 bc_1 复合体）和复合体Ⅳ（细胞色素 c 氧化酶）泵出膜从而形成了跨膜的质子梯度差。这个梯度差驱动 ATP 合酶合成 ATP。复合体 II（琥珀酸脱氢酶）没有参与泵出质子。

5.3.2.1　化学渗透理论

合成 ATP 的过程被称为氧化磷酸化（oxidative phosphorylation）：



$$ADP+P_i \rightleftharpoons ATP+H_2O,\quad \Delta_r G^{\ominus}=+31kJ\cdot mol^{-1}$$

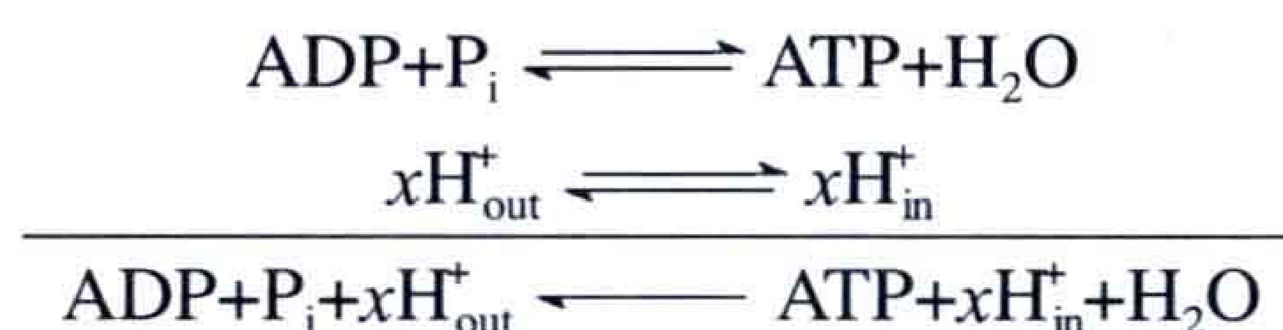
合成 ATP 的反应使得细胞中的 ATP 浓度远远大于上述平衡中的产物浓度。而这就被细胞中许多需要 ATP 参与来完成其功能的生物学过程利用。我们知道 ATP 在线粒体中是直接由 ADP 和磷酸合成的。许多年来人们都不知道这是如何发生的。在 1961 年，皮特·米歇尔（Peter Mitchell, 1978 年诺贝尔化学奖获得者，见附录 F）提出了以下几个必要的特点。

• 能量转换的（与氧化和 ATP 生成相关的）中间体是跨膜的离子梯度差。

• 参与形成梯度差的离子是质子（或者 H_3O^+）。

以上两个方面构成了化学渗透理论的基本假设。“化学渗透（chemiosmotic）”这个名词表明了化学反应和储存于跨膜梯度差中的能量（“渗透能”）的关联。因此，根据米歇尔的理论，电子在通过传递链转移的过程中所释放的能量与氢离子浓度梯度差和电势是对应的，并且这个梯度差驱动了氧化磷酸化。伴随着电子顺电子转运链“流下”，氢离子被从线粒体的膜内驱赶至膜间（图 5.11）。这导致了线粒体内膜之内的 pH 上升而之外的 pH 下降——于是一个持续的 pH 梯度就这样产生。而跨膜的电势差也同时上升，因为膜外比膜内有更多的带正电荷的氢离子。建立 pH 梯度和电势差这两个过程是需要消耗吉布斯自由能的。米歇尔称其为质子动力，我们将对此详细讨论。

在人类及其他动物中，质子动力产生在线粒体的内膜两侧，而在植物中则产生在叶绿体的内膜两侧，在需氧菌中则发生在质膜两侧。因此，化学渗透的机制取决于跨膜的 pH 梯度，而这种梯度是通过一系列的电子传递建立的。ATP 的合成同时伴随着质子的跨膜传递。总体的反应可以写成如下两个反应方程的简单加和：



$$ADP+P_i \rightleftharpoons ATP+H_2O$$
$$xH^+_{out} \rightleftharpoons xH^+_{in}$$
$$ADP+P_i+xH^+_{out} \longleftarrow ATP+xH^+_{in}+H_2O$$

跨膜氢离子梯度所储存的能量来自于两方面。第一个方面来自于膜两侧不同的氢离子活度导致的不同的化学势：

$$\Delta \mathbf{G}_{m}=G_{m,in}-G_{m,out}=RT\ln(a^{+}_{H,in}/a^{+}_{H,out})$$

这种化学势来源于膜两侧的混合物有着不同的熵值。

第二个方面来自于膜两侧不同的静电相互作用（膜外为正、膜内为负）而产生的膜电势差 $\Delta\Phi=\Phi_{in}-\Phi_{out}$。每一摩尔的氢离子所产生的跨膜电荷差为 $N_{A}e=F$，即法拉第常数，而这个过程的 ΔG 等于 $F\Delta\phi$。

因此，当一个质子从膜内被转移到膜外，总共储存的吉布斯能为

$$\Delta G_{trans}=F\Delta\phi-(R\ln 10)\Delta pH$$

在此我们用活度代替摩尔浓度并且引入 $pH=-\log[H^{+}]$ 及代入下式，即

$$\Delta pH=pH_{in}-pH_{out}=-\log[H^{+}]_{in}\pm\log[H^{+}]_{out}$$

在线粒体中 $\Delta pH\approx-1.4$，对应于温度 25℃时的 ΔG 为 8 kJ · mol^{-1}；而 $\Delta\Phi\approx0.14$ V，对应于 25℃时的 ΔG 为 13.55 kJ · mol^{-1}，于是 25℃时的 $\Delta G_{trans}=21.5$ kJ · mol^{-1}。由于 ADP 的磷酸化需要 31 kJ · mol^{-1} 的能量，所以至少需要跨膜转运 2 mol 的氢离子（可能更多，因为我们并没有考虑被动的质子穿膜泄露）以磷酸化 1 mol 的 ADP。线粒体的有效电容量及缓冲液容积决定了自由能中 $\Delta\phi$ 和 ΔpH 的分配。

5.3.2.2 结构

ATP 合酶第一次作为明确的跨线粒体膜结构被观察到是通过电子显微镜（EM）手段（图 5.12）。接下来，这些结构被鉴定为正在合成 ATP 的结构。

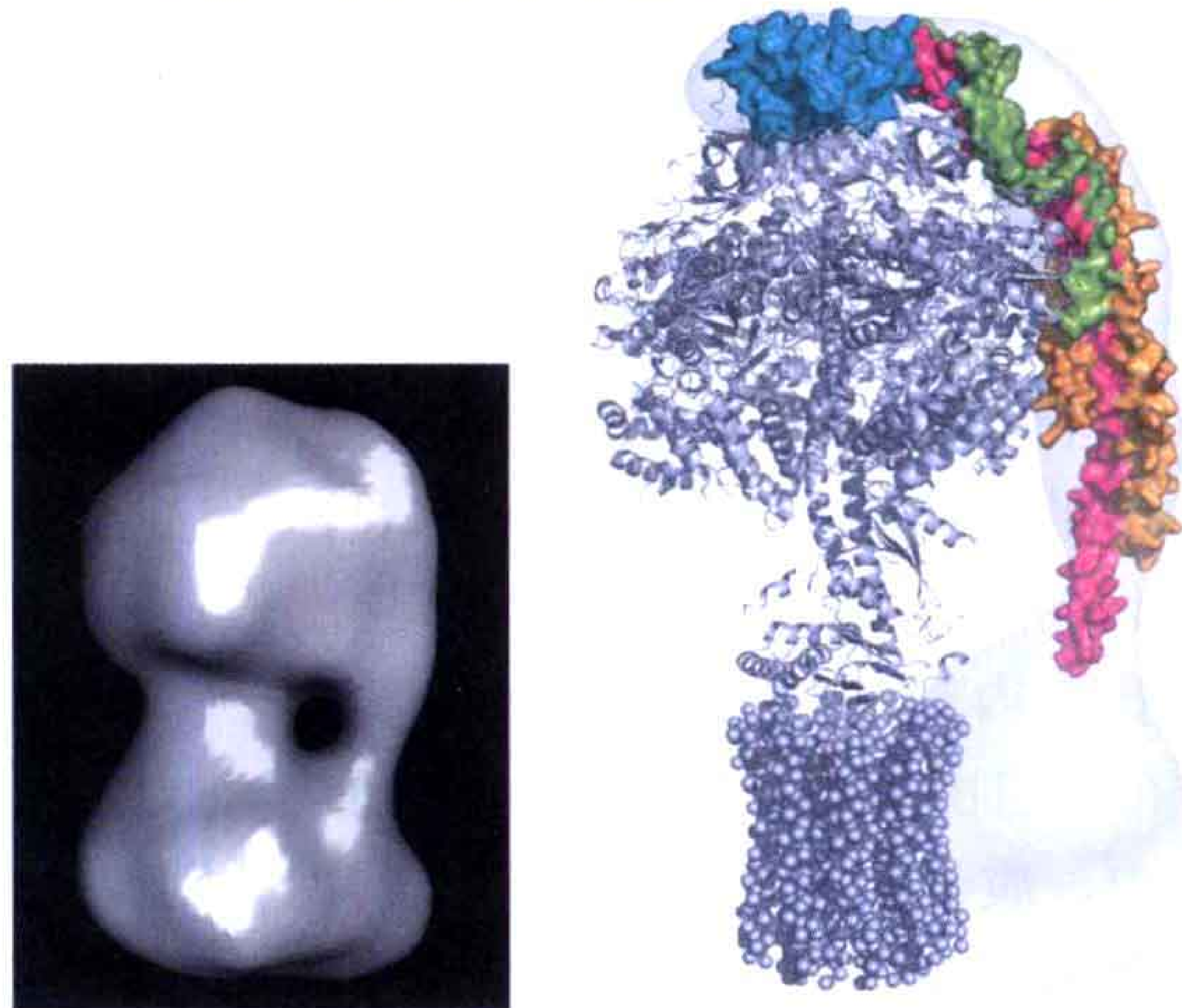


图 5.12 左：电镜观察到的线粒体 ATP 合酶结构。图中主要显示全酶的两个大组件，即上面的 F_1 和下面的 F_o。两个组件之间除了中心有柄连接外，右侧还有一个较弱的次要的柄 [获刊印授权，原文出自 Rubinstein, Walker, & Henderson. (2003) Structure of the mitochondrial ATP synthase by electron cryomicroscopy. *EMBO J* **22**:6182-6192. 版权所有 Nature Publishing Group]。右：ATP 合酶分子的晶体结构吻合入电镜密度。F_1 部分及 c 亚基为灰色，而其余部分及侧柄显示为不同颜色（获刊印授权，原文出自 Dickson, Silvester, Fearnley, Leslie & Walker. (2006) On the structure of the stator of the mitochondrial ATP synthase. *EMBO J* **25**:2911-2918。版权所有 Nature Publishing Group)。

ATP 合酶有两个独立元件，即 F_o 及 F_1。F_o 为穿膜结构而 F_1 位于膜的胞质一侧。两个元件中的亚基分别属于马达中的转子和定子（表 5.4）。在其行使功能的过程中，跨膜的质子梯度驱使 ATP 合酶的转子转动，而定子部分仍然是固定的（图 5.13）。

表 5.4　ATP 合酶的亚基

亚基	化学计量			
	线粒体	叶绿体	细菌	转子 / 定子
F_o				
a	1	1	1	定子
b	1	b + b′	2	定子
c	10 ~ 15	10 ~ 15	10 ~ 15	转子
d	1	—	—	定子
F_6(h)	1	—	—	定子
OSCP[a]	1	δ	δ	定子
F_1				
α	3	3	3	定子
β	3	3	3	定子
γ	1	1	1	转子
δ[b]	1	—	—	转子
ε	1	1	1	转子

a 在细菌及叶绿体中称为 δ。
b 在细菌及叶绿体中称为 ε。

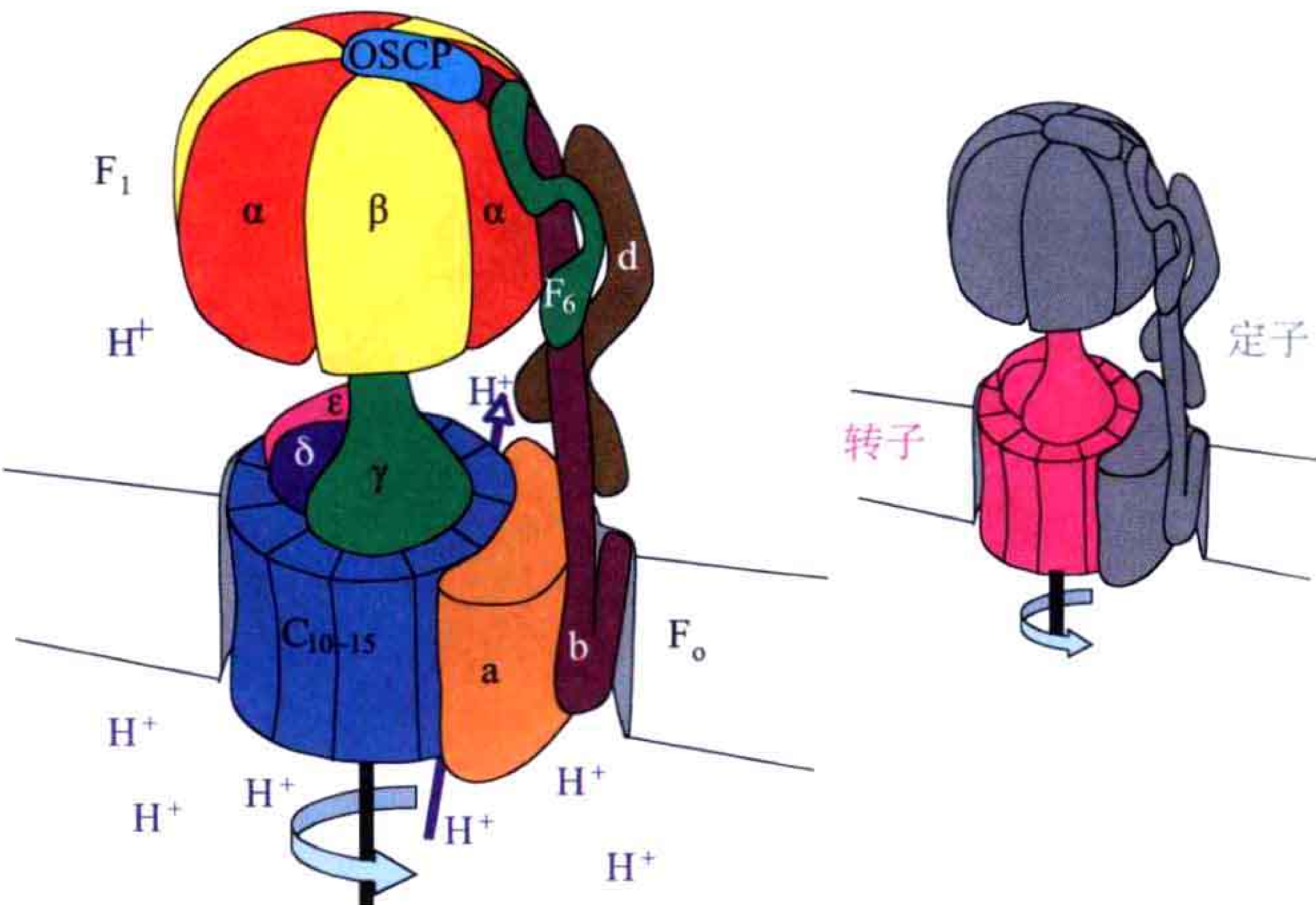


图 5.13　ATP 合酶的组成形式。跨膜的质子梯度驱使复合体的其中一部分（转子）随着质子的跨膜转移而转动，同时生成 ATP。F_o 部分与膜相连而 F_1 部分伸出膜外。左：F_o 的环状 c 亚基（蓝色）在膜中转动，而 F_1 的 γ 亚基（绿色）、δ 亚基（深蓝色）和 ε 亚基（桃红色）也随之转动。F_o 的 a 亚基（橙色）、b 亚基（深紫色）、d 亚基（棕色）、F_6（绿色）和 OSCP 亚基（蓝色）与 F_1 的三个 α 亚基（红色）和三个 β 亚基（黄色）组成了定子。右：ATP 合酶的定子部分显示为灰色，而转子部分显示为红色。

F_1 由三个重复单元组成，每个单元由两个亚基组成，这两个亚基称为 α 和 β。除三个重复单元外 F_1 还包含一个 γ、一个 δ 和一个 ε 亚基。ATP 合酶的活性位点就在三个 β 亚基上。三个 β 亚基与三个 α 亚基（与 β 亚基十分相似）相间排列成一个六聚体。

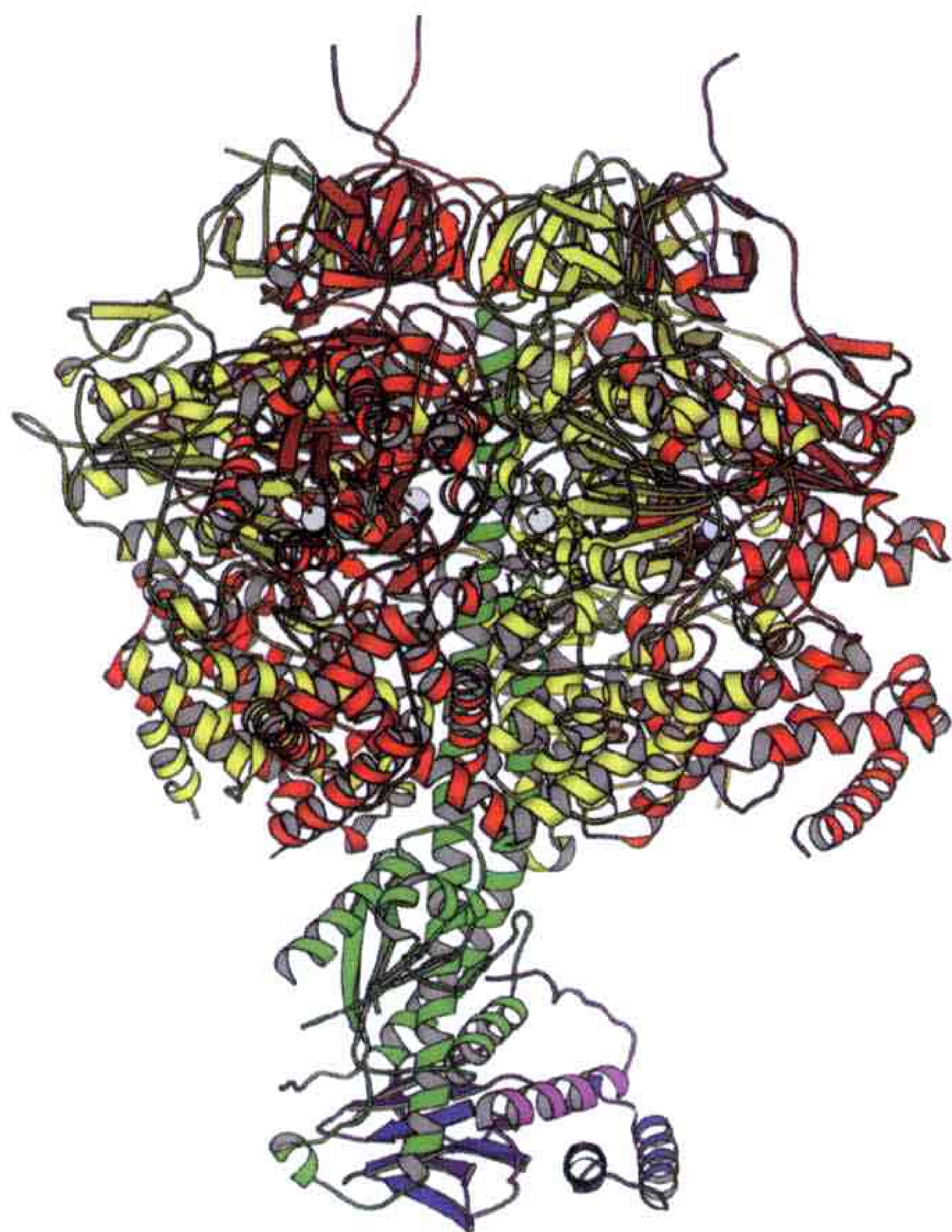


图 5.14　ATP 合酶 F_1 部分的结构。绿色所示为 γ 亚基，γ 亚基作为一个柄插入具有三重对称性质的 $α_3β_3$ 聚合体中。γ 亚基本身是没有对称性的，所以可以与三个催化 β 亚基形成不同的相互作用（PDB：1E79）。

γ 亚基中较细的部分贯穿这个六聚体。这一部分是一个左手 α 螺旋的卷曲螺旋（coiled coil）。而 δ 和 ε 亚基依附在 γ 亚基较宽阔的底座上（图 5.13）。由于其不对称性，γ 亚基与三个 β 亚基以不同的方式接触，并使得三个 β 亚基处于不同的构象（图 5.14 和图 5.15）。

α 亚基和 γ 亚基都有三个结构域：一个 N 端的 β 桶结构位于靠近整个分子顶端的位置（图 5.15），一个结合核苷酸的催化结构域位于中部，以及一个 C 端的螺旋结构域位于靠近膜的位置。催化位点位于 α 亚基和 β 亚基的界面上。ATP 合酶属于一个叫 AAA+ ATPase 的酶家族。ATP 合酶的 N 端结构域形成一个假（pseudo）6 次轴对称的环。这个环为 $α_3β_3$ 复合体提供了很强的稳定性。α 亚基和 β 亚基的中心结构域有着与 RecA 蛋白相同的折叠（见表 5.3）。

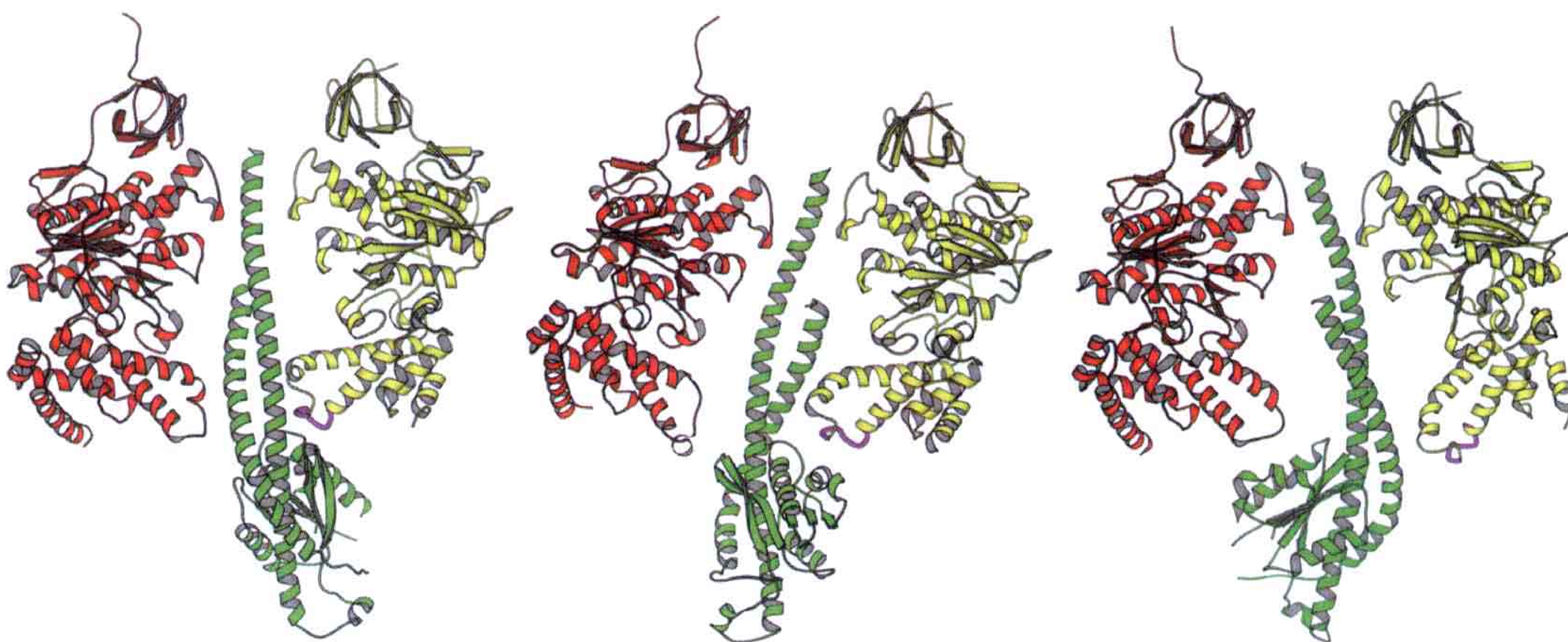


图 5.15　三个图分别所示为 γ 亚基与三组位置相对的 α 亚基和 β 亚基相互作用的情况。由图可见不对称的 γ 亚基与三个 α 亚基（红色）和三个 β 亚基（黄色）有着不同的相互作用。可以观察到三个 β 亚基上的活性位点的构象各不相同，可分别称为空状态 $β_E$（empty）、结合 ADP 的状态 $β_{DP}$（with a bound ADP）及结合 ATP 的状态 $β_{TP}$（with a bound ATP）。C 端的一段环（紫色）与 γ 亚基的卷曲螺旋部分相互作用并且影响了 β 亚基上的活性位点的构象。对于结合了 ADP 和氟化铝的 ATP 合酶来说，三个活性位点虽然仍处于不对称的状态，但都结合了 ADP。其中 $β_{DP}$ 和 $β_{TP}$ 结合位点包含着 ADP*AlF_4^- 且处于一个关闭的构象，而对于 $β_E$、ADP 和硫酸根离子是以一种半关闭的构象结合的，这种构象又被称为 $β_{HC}$。

不同的生物组织中的 Fo 或多或少有一些差异。在细菌和叶绿体中，F_o 是由 a 亚基、b_2 亚基、$c_{10\sim15}$ 亚基组成（表 5.4）。在线粒体中 Fo 同样也有 d 亚基和 F_6 亚基，但是只有一个拷贝的 b 亚基。d 亚基和 F6 亚基部分地代替了缺失的那个拷贝的 b 亚基的功能。线粒体中那个单独的 b 亚基在 N 端含有两个跨膜的螺旋。在细菌中的 δ 亚基对应于线粒体中的 OSCP 亚基。b 亚基、d 亚基、F_6 亚基和 OSCP 亚基均属于马达中的次要的，或者说是属于定子的那个柄。OSCP 亚基（在细菌和叶绿体中是 δ 亚基）黏附在 F_1 的一个或一些 α 亚基的 N 端，以防止 α 和 β 亚基转动。

在 β 亚基的 C 端结构域，有一个处于两个螺旋之间的环与 γ 亚基的卷曲螺旋部分发生相互作用（图 5.15）。这影响了 β 亚基的构象及活性位点的状态。这个环的一部分含有 DELSEED 的一级序列。其中前两个残基（394 位的天冬氨酸和 395 位的谷氨酸）与 γ 亚基上相应部分的大量正电荷相互作用（图 5.16）。而这些残基的其中一些是高度保守的。γ 亚基上的这些正电荷构成了一个离子的轨道，这个离子轨道与 β 亚基上那段环的酸性残基相互作用，从而在 γ 亚基转动时为这些酸性残基的负电荷提供了一个低能量的路径。在转动过程中 β 亚基和 γ 亚基所经历的空间位阻及静电相互作用最终实现了 ATP 的合成，反过来，其在 ATP 水解时也可以驱动这个转动。

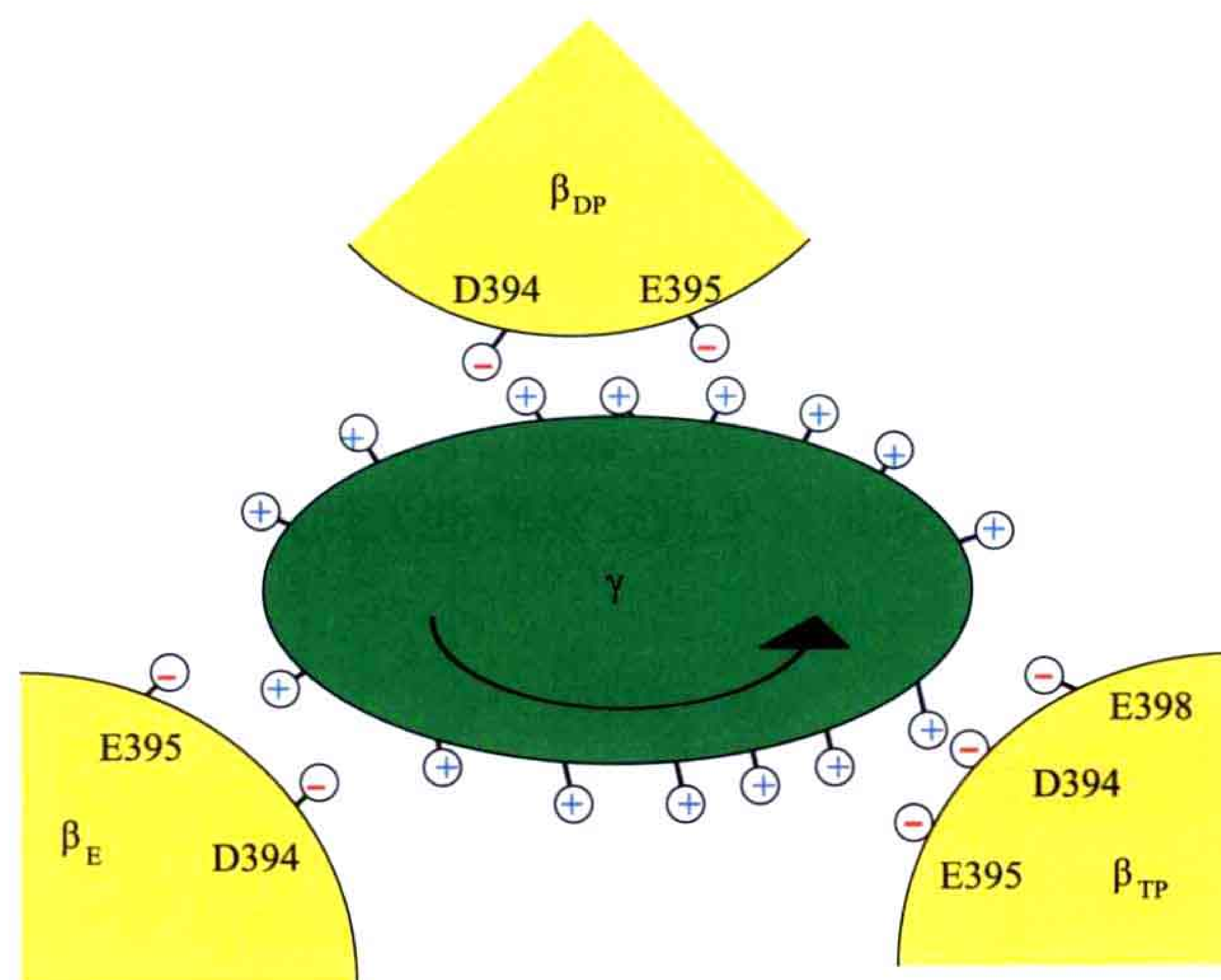


图 5.16 γ 亚基的卷曲螺旋部分（绿色）含有一定数量的带正电荷的残基，从而可以和 β 亚基（黄色）上的负电荷相互作用。

5.3.2.3 ATP 水解和 ATP 合成

ATP 合酶的 F_1 部分同样可以作为 ATP 水解酶使用。根据催化的机理，Paul Boyer 计算了当 ATP 水解时三个活性位点必须经历有效的构象变化。而全酶在利用质子移动的力量合成 ATP 时必须经历类似的构象变换（图 5.17）。当得到 F_1 的晶体结构以后这个“结合变换机理”得到了部分的证实，即 γ 亚基的不对称性驱使了 β 亚基的不同构象（图 5.15）。最初被讨论到的三种 β 亚基的构象为：开放（open）、松弛（loose）和紧密（tight）。

在接下来得到的结合了 ADP 和氟化铝的晶体结构中，三个活性位点虽然仍保持了不对称性，但都结合了核苷酸。其中 β_{DP} 和 β_{TP} 结合位点包含着 ADP*AlF_4 且处于一个关闭的构象，而 β_E 位点中结合着一个 ADP 分子及一个硫酸根离子，而且是以一种半关闭的构象结合的，这种构象又被称为 β_{HC}。

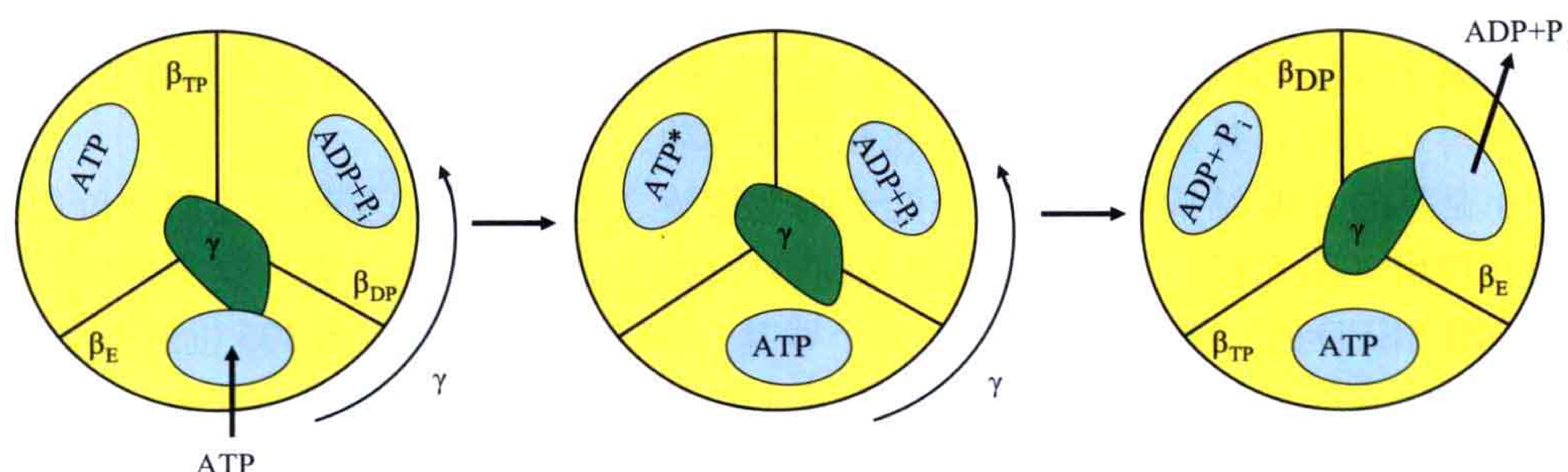


图 5.17　ATP 合酶的水解机理。这个机理最初由 Boyer 提出，之后根据晶体结构修正。图中可见三个 β 亚基在行使功能的循环中分别处于不同的构象。其中 β_E 亚基是空的状态，β_{DP} 包含了一个已经被水解的 ATP 分子，而 β_{TP} 则结合了一个 ATP 分子。当 ATP 结合到 β_E 时，γ 亚基就 β 亚基而言发生了转动。这致使之前的 β_{TP} 亚基中的 ATP 发生活化。当 ADP 和无机磷酸（P_i）被释放，γ 亚基将继续转动并导致了 β_E 又处于开放构象。ATP 合成的过程则沿反方向进行，驱动力是跨膜的质子梯度。

关于转动特性的最终确认是通过如图 5.18 所示的实验完成的。β 亚基的 N 端被加上了组氨酸标签（His-tag）并且被绑定在包被了镍的盖玻片上。γ 亚基则被黏上了一段荧光标记的肌动蛋白丝。在显微镜下这段肌动蛋白丝显示为一个荧光的小杆。当加入 ATP 时，这个小杆以逆时针转动（从膜上看的方向），这使得 ATP 合酶发生的分子事件可以被观测到。当进一步分析时，发现这种转动大约每 30° 或 90° 一步。当转动

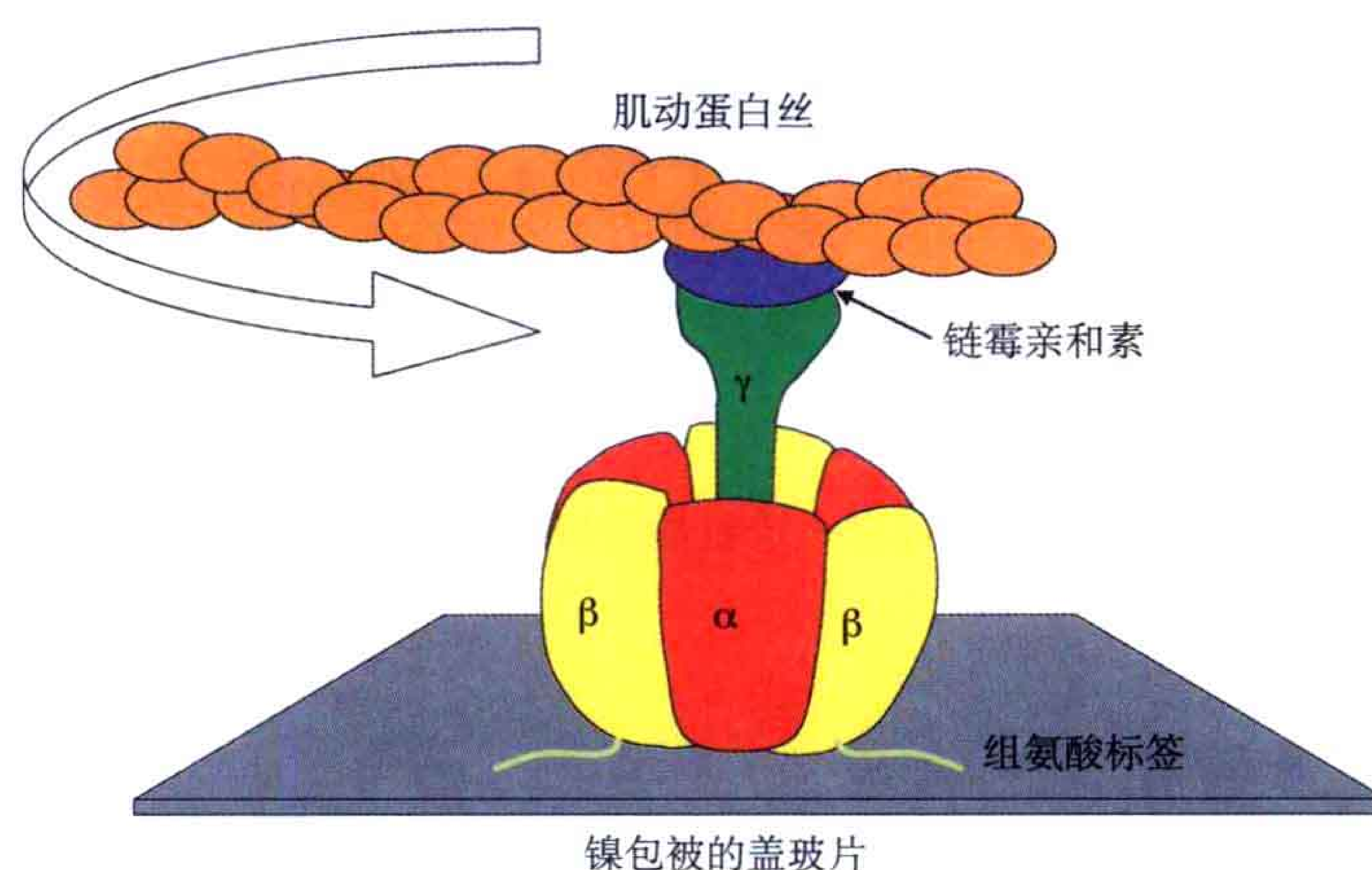


图 5.18　转动理论的证据是这样一个实验：ATP 合酶的 β 亚基 N 端被加上一段组氨酸标签，这个标签对镍有很强的亲合力，于是 F_1 就被这样黏在一个包被了镍的盖玻片上。而柄的转动是通过观察与之相连的一段荧光的肌动蛋白丝的转动。这个肌动蛋白丝是通过一个叫链霉素的蛋白质结合在 γ 亚基上的。链霉亲和素与肌动蛋白丝上的生物素及 γ 亚基都有很高的亲合性。加入 ATP 后，这段肌动蛋白丝可以以逆时针方向转动，转动的速率取决于 ATP 的浓度。

120° 后，一个 ATP 分子就被消耗。总体上看 β 亚基都处于与 30° 旋转前相同的状态，但是每一个单独的 β 亚基在转动 360° 后都向前前进了一个步骤。除了水解 ATP 并旋转之外，最近的研究发现如果用外力驱使 γ 亚基以顺时针转动，那么单独的 F_1 部分也可以合成 ATP。

ATP 合酶的催化机理与其他 ATP 水解酶和 GTP 水解酶是相同的。它们都有一个 P-loop，这个 P-loop 含有一个高度保守的 GGAGVGKT 一级序列，而这个序列可以与核苷酸相互作用。图 5.19 显示的是位于 β 亚基上的活性位点。

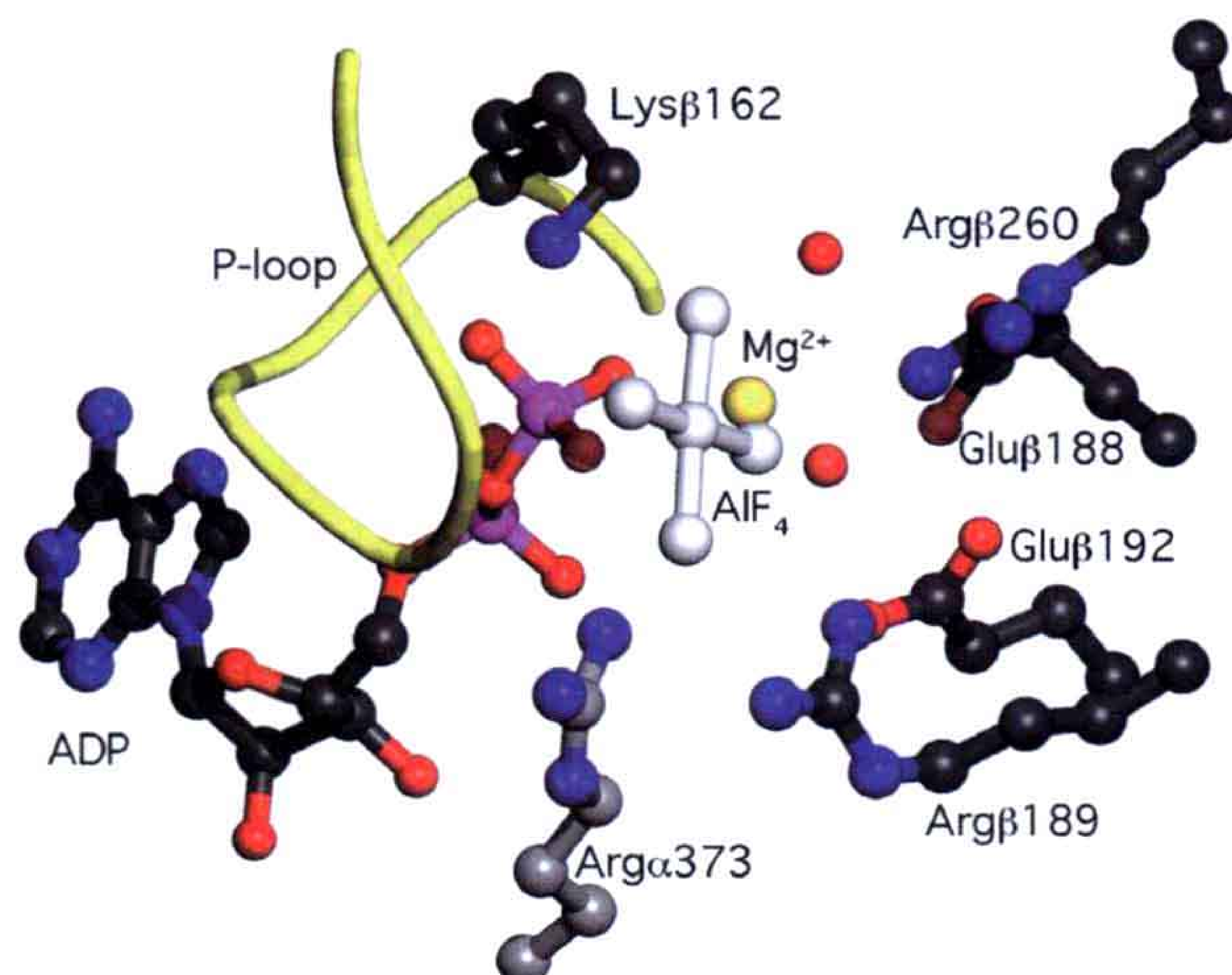


图 5.19 处于 β_{DP} 构象的 β 亚基的活性位点。用来模拟 ATP 的是一个 ADP 分子和一个氟化铝离子，氟化铝所结合的位置也就是 γ 磷酸的位置。这个“ATP”分子与 P-loop 上第 162 位的赖氨酸及一些精氨酸相互作用。α 亚基则贡献了第 373 位的精氨酸（碳原子以灰色显示），这个精氨酸扮演的角色与 GTP 水解酶中的“精氨酸指（arginine finger）”一致。188 位的谷氨酸对应于 GTP 水解酶中参与激活水分子以水解 γ 磷酸的那些残基。Lys：赖氨酸；Arg：精氨酸；Glu：谷氨酸。

三个 β 亚基对于 ATP 和 ADP 的亲合性对于了解这个机理是十分重要的，尤其是这个酶如何在 ATP 可以自然水解的条件下合成 ATP 及如何避免 ATP 对酶的抑制作用？讨论这两个问题可以使用两个机理，即双位点机理（bisite mechanism）和三位点机理（trisite mechanism）。在双位点机理中，在特定时间只有两个位点在运行而另一个则是空着的。而在三位点机理中，所有的三个位点可以被同时占据。ATP 合酶看似是以三位点机理运转的，ATP 和 ADP 的负协同作用对于产生驱动 γ 亚基转动的力量来说是必需的。这个机理相对于单位点催化机理有着很大的先进性。

一共有 4 种不同的 ATP 结合常数：0.2nmol/L、2μmol/L、25μmol/L 和 5mmol/L。β_{TP} 结合 ATP 或者 ADP + P_i 的自由能大致是相同的，而 β_{DP} 则非常倾向于结合 ADP + P_i。在水解反应时，ATP 结合在开放的 β_E 位点上，而当位点由 β_E 转变为 β_{TP} 状态时，则可获得大量的结合自由能。当另一个 ATP 分子结合在一个新的 β_E 位点时，则驱动了正在 β_{TP} 状态的位点转变为 β_{DP} 状态。于是由于 β_{DP} 相较于结合 ATP 更倾向于结合 ADP+ P_i，所以 β_{DP} 中的 ATP 分子被水解成 ADP + P_i。这些构象变化驱动了 γ 亚基相对于 α 和 β 亚基的转动。

ATP 的合成是被 γ 亚基的顺时针转动驱动的。处于半关闭状态即 β_{HC} 状态的 β 亚基对于 ADP 的亲合性小于对 ATP 的亲合性。γ 亚基的转动使得 β 亚基变为 β_{DP} 状态，而在下一个转动步骤则变成 β_{TP} 状态。此时结合 ADP + P_i 和结合 ATP 的结合自由能是相同的，而缓慢的 ATP 合成则正在发生。快速的 ATP 合成应该是基于一个我们迄今为止没有发现的状态。在 γ 亚基发生进一步转动之后，我们又重新回到了 β_E 的状态，这时位点是开放的且对 ATP 只有很低的亲合性。于是 ATP 就被游离出来。这个过程也不会被高浓度的 ATP 所抑制。

如图 5.19 所示为牛的线粒体 ATP 合酶中一些在催化机理中非常关键的残基。162 位赖氨酸便是处于 P-loop 上的赖氨酸，这个赖氨酸在所有 ATP 水解酶和 GTP 水解酶家族中都是保守的。188 位谷氨酸作用于参与水解反应的那个水分子，或者是在合成反应时作为产物的水分子。这个残基对应于许多 GTP 水解酶中的谷氨酸及用于蛋白质翻译的 GTP 水解酶中的一个组氨酸。189 位的精氨酸残基与 γ 磷酸形成了氢键，它在 GTP 水解酶中并没有对应的残基。在活性位点的所有有功能的残基中唯一一个来自于 α 亚基的是 373 位的精氨酸，它对应于 GTP 水解酶中的“精氨酸指（arginine finger）”。这两个精氨酸在不同的状态中处于不同的位置，这对不同状态对核苷酸不同的亲合性有很重要的影响，从而影响了反应的发生。

5.3.2.4　一个马达酶

质子运动的力量如何能使得 γ 亚基转动以合成 ATP？跨膜的 pH 梯度运作了 F_o 部分。F_1 的 γ 亚基和 ε 亚基结合在环状的 c 亚基上。c 亚基的数量在不同 ATP 合酶中是不同的。在酵母 ATP 合酶中的数量是 10 个，而 11 个［来自酒石酸泥杆菌（*Ilyobacter tartaricus*）ATP 合酶；图 5.20］一直到 15 个的情况都有被发现。因此 c 亚基环的对称性常常不与 F_1 上三重轴的 α 和 β 亚基相一致。ATP 合酶能够在一个较低的离子运动力（ion-motive force）中运作就是得益于大数量的 c 亚基。含有越少 c 亚基的 ATP 合酶其离子运动力越大。

c 亚基由一个横穿膜的螺旋发夹结构组成。参与结合质子的残基是一个在中途穿出膜结构的天冬氨酸。图 5.20 显示的状况是钠离子的梯度在驱动这个马达。钠离子结合在膜的中部，跨越 c 亚基的一部分及亚基的交界。钠离子的主要配体是一个谷氨酸、一个谷氨酰胺和一个丝氨酸。其中谷氨酸对应于质子结合时的天冬氨酸。

a 亚基形成了一个供质子或离子通过的通道。a 亚基的结构目前尚不清楚，但是可以想像这个通道是由两个不相连接的“半通道”组成的（图 5.21）。这个模型被称为“双通道模型”。当一个质子或者钠离子通过入口的通道从有较高离子浓度的一侧进入时，c 亚基的质子或者离子结合位点就会被装载。然后 c 亚基被引导转动几乎一整圈以便将这个质子或离子通过出口通道释放到低离子浓度的那一侧去。当转动的 c 亚基靠近属于定子的 a 亚基时（图 5.22），一个定子上的精氨酸与 c 亚基上固定钠离子（质子）的谷氨酸（天冬氨酸）相互作用。于是这个离子被释放到出口通道而一个新的离子则被入口通道获取。这个机理可以被描述为“推与拉（push-and-pull）”机理，a 亚基的

精氨酸与质子或钠离子竞争与酸性侧链相互作用。

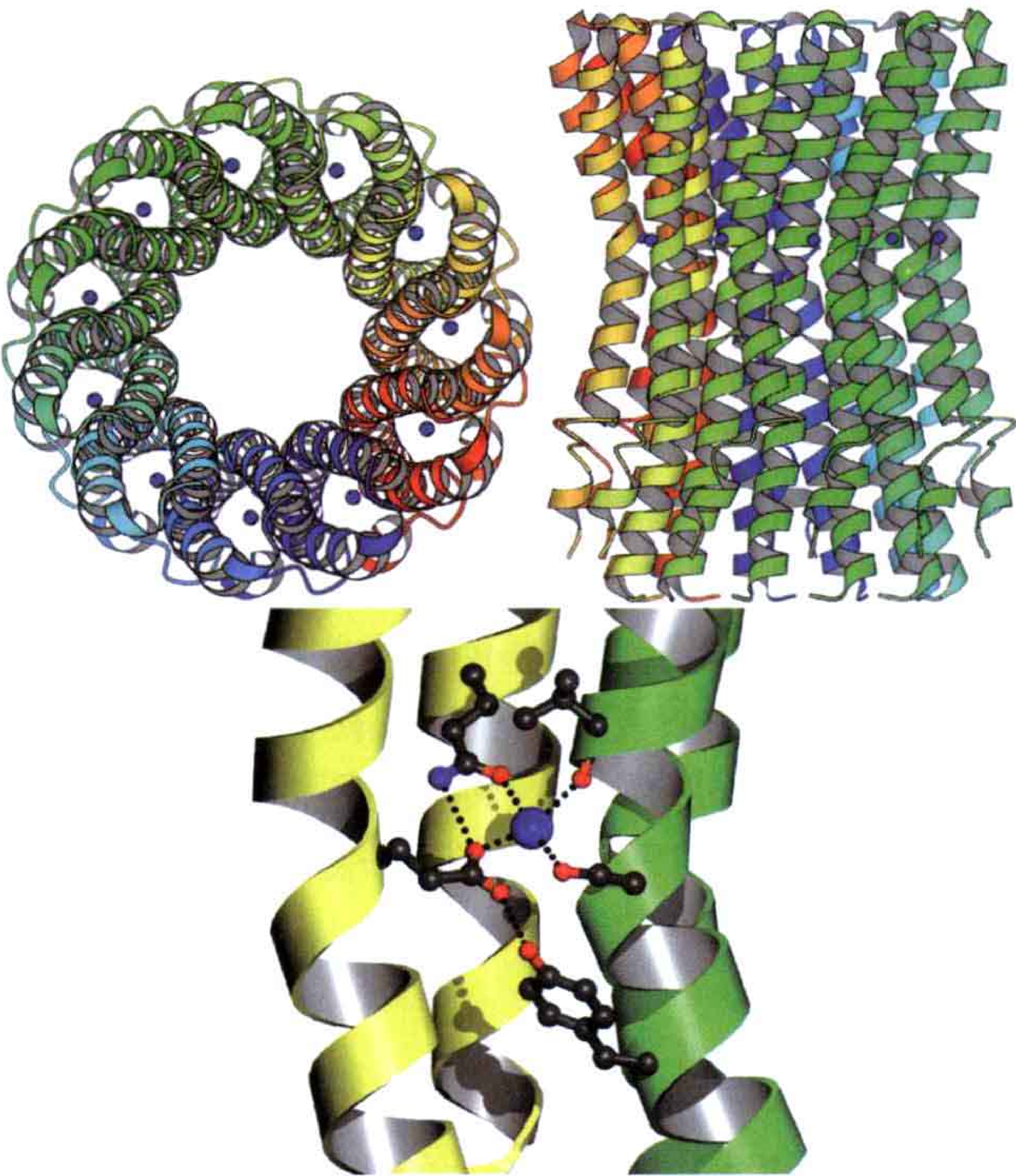


图 5.20　被钠离子驱动的来自酒石酸泥杆菌的 ATP 合酶的结构。这里 11 个 c 亚基形成一个环并穿过膜。上：分别为垂直于膜和切面的视角；下：所示为一个钠离子结合在两个亚基（黄色和绿色）之间（PDB：1YCE）。

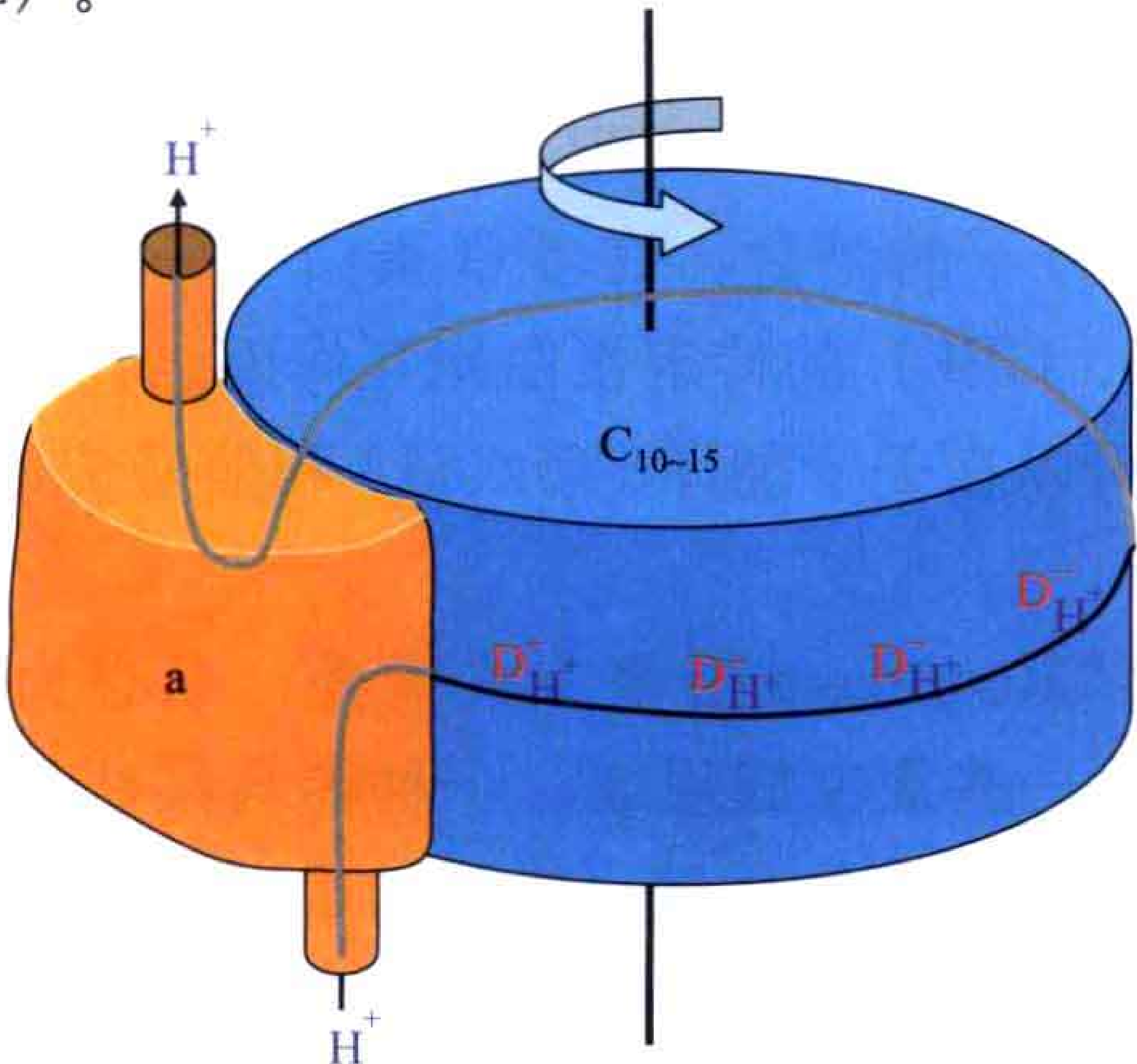


图 5.21　Fo 中亚基的排列情况，其中 c 亚基被用来在膜中转动。a 亚基形成一个半通道以使质子（或钠离子）通过并到达与之结合的 c 亚基的天冬氨酸处。为了释放这个质子（或者钠离子），c 亚基以棘齿（ratchet）的原理转动至一个新的位置，在这个位置有 a 亚基的另一个半通道。这种工作方式有点类似于使用水车的磨坊，即以流水驱动轮子转动以做功。

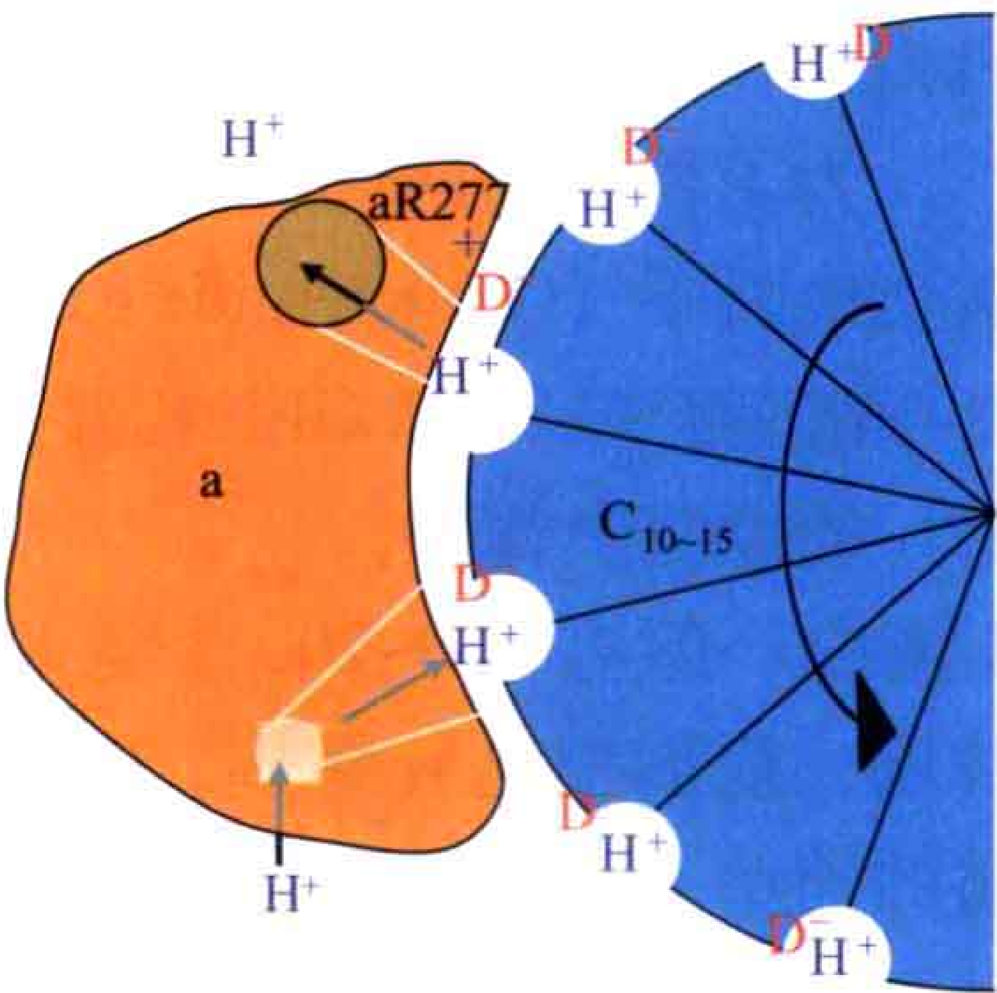


图 5.22　ATP 合酶的“推与拉”机理。c 亚基（蓝色）的天冬氨酸侧链暂时与属于定子的 a 亚基（橙色）上的一个精氨酸相互作用。这使得天冬氨酸失去了本与之结合的质子并将质子通过出口通道（图上方）释放。而一个新的质子将通过入口通道（图下方）结合在下一个 c 亚基的带负电荷的天冬氨酸上。因此质子的通过可以驱动这个 c 亚基的“轮子”转动并合成 ATP。

ATP 的合成不仅需要质子或钠离子的梯度，同时还需要膜电位。在放松的模式下，这个马达可以向两个方向转动并且执行膜两侧的离子交换。而当膜电位和质子或钠离子梯度都存在时，酶便转变成只能向一个方向转动。

c 亚基的转动也涉及 γ 亚基、δ 亚基和 ε 亚基。这些亚基构成了马达酶的转子部分。其中 γ 亚基形成了酶的中心柄或者说是转动柄，而次要柄则连接 F_o 中的定子部件和 F_1 部分。定子包括 F_o 的 a、b、d、F_6 和 OSCP 亚基及 F_1 的 α 和 β 亚基。这个次要柄在冷冻电镜（cryo-EM）的研究中可以被观察到。定子中一部分的结构已经通过晶体学解析并且可以与冷冻电镜的密度相符合。其中一个结构显示 b 亚基的一部分为一个延伸的螺旋，由 105 个连续的残基组成。它组成了整个柄的骨架。d 亚基有 5 个螺旋，其中 3 个沿着 b 亚基的螺旋排布并对其提供支撑。F_6 亚基包含位于 b 螺旋上方的两个螺旋。

c 亚基的环需要转动 10~15 步以转完一整圈。由此算来合成一个 ATP 分子需要 3 ~ 5 个质子，因为转一整圈将生成 3 个 ATP 分子。显然转动会产生一个作用在中心柄和次要柄上的扭力。这个扭力在不同的物种及转动的不同步骤中均有不同。而扭力的释放出现在两个确定的步骤，即 30° 和 90° ，相当于一整圈的 1/3 及一个 ATP 分子的生成。

5.3.3 G 蛋白

G 蛋白（或者说是 GTP 水解酶）是一个庞大的酶家族。所有的 GTP 结合结构域（或者称为 G 结构域）通常都有类似的折叠方式并以同样的方法结合核苷酸。G 蛋白是一种分子开关，其有一个开启（ON）状态和一个关闭（OFF）状态。GTP 水解成 GDP 和无机磷酸（P_i）的反应通常是在 G 蛋白与其他蛋白质相互作用后引发的，而这种与 G 蛋白相互作用的蛋白质就被称为 GTP 水解酶激活蛋白（GTP ase activating protein，GAP）。为了催化 GDP 的解离及促进新 GTP 分子的结合，许多 G 蛋白与一种称为 G 核苷酸交换因子（G nucleotide exchange factor，GEF）的特定蛋白相互作用。其中一些 G 蛋白与相应的 GAP 和 GEF 相互作用的例子可以在关于翻译（第八章）和信号传递（第十一章）的章节中找到。

G 蛋白一般有 160~200 个氨基酸残基并以 Rossmann 折叠方式折叠，中心为平行的 β 片层，而周围是 α 螺旋。G 蛋白的家族由一些亚家族组成。一种参与某些信号传导通路的简单 G 蛋白 Ras 仅由一个结构域组成，即 G 结构域（详见第十一章）。其他的成员则是多结构域的蛋白质，其中包括 G 结构域。另一个亚家族是更加复杂的异源三聚体 G 蛋白，参与其他的一些信号通路。G 蛋白在结构上和功能上均属于一个大的 ATP 水解酶家族。

5.3.3.1 结构、一致性元件、核苷酸及镁离子结合

Ras 蛋白中心是由大部分平行的 β 片层组成，两侧则是螺旋结构（图 5.23）。这与在许多结合核苷酸的酶中发现的 Rossmann 折叠十分相似。但不同之处在于 Ras 蛋白

的 β 股顺序为 231456，其中 β 股 2 是反平行于其他股的（Rossmann 折叠的 β 股顺序为 321456 且所有的 β 股都是平行的）。

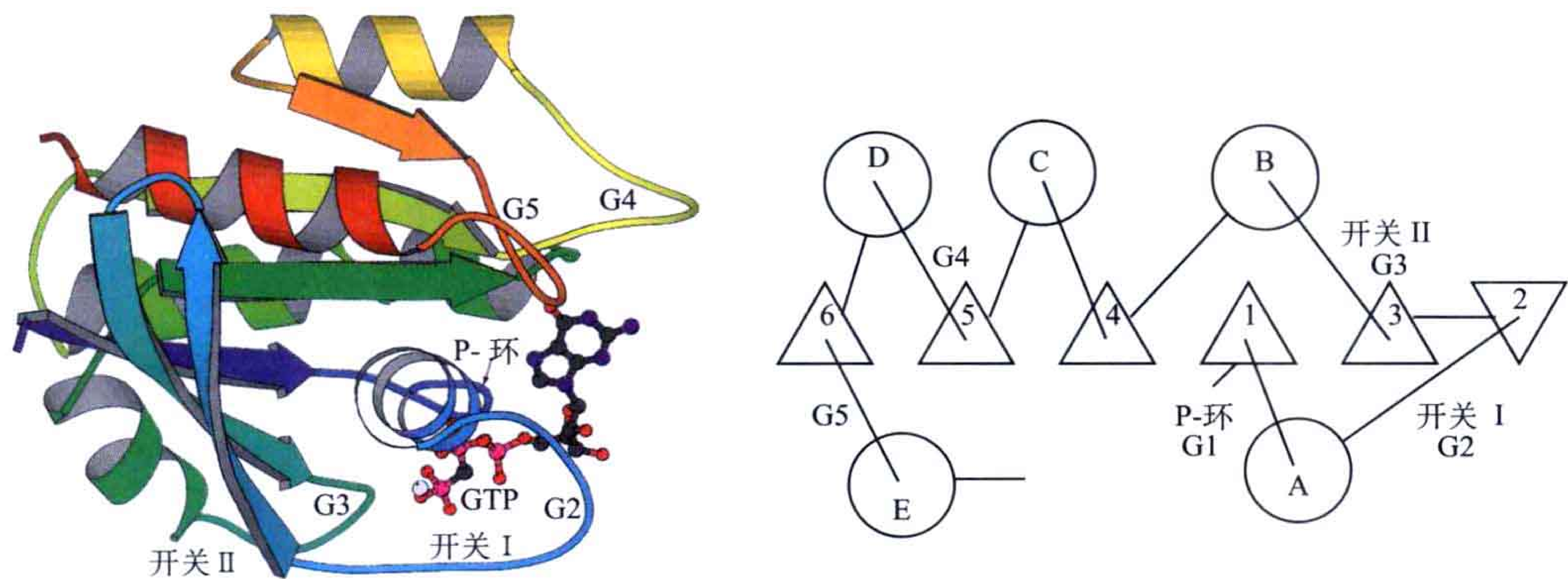


图 5.23　左：G 结构域——此图中为 Ras 蛋白的 G 结构域，Ras 蛋白活跃在许多信号转导通路中。着色的顺序是从 N 端（蓝色）到 C 端（红色）（PDB：121P）。右：G 结构域的组成。圆圈代表 α 螺旋，三角代表 β 股。每个二级元件上的数字分别代表了这个元件的序号。5 个含有一致元件的 loop 被高亮显示。P- 环、开关 I 和开关 II 对于与磷酸相互作用及识别 γ 磷酸来说是必需的。G4 和 G5 loop 用来鉴别 G 核苷酸。

G 结构域的核心部分在所有 G 蛋白中都是保守的，但是整个结构域中的变化是常见的。G 结构域的特点是含有 5 个一致性元件或者模体（consensus elements or motif）（G1 ~ G5）（表 5.5）。

表 5.5　G 蛋白的一致元件

元件	名称	序列	功能
P- 环	G1	GXXXGKT/S	与 α 和 β 磷酸相互作用
开关 I	G2	XTX	结合 γ 磷酸和镁离子
开关 II	G3	DXXG	结合 γ 磷酸并间接结合镁离子
—	G4	N/TKXD	识别 G 核苷酸
—	G5	T/GC/SAL/K	结合 G 核苷酸

5 个一致元件主要位于 β 股的 C 端和 α 螺旋的连接处并且组成了核苷酸和镁离子的结合位点（图 5.24）。总的说来，前三个一致元件（P-loop、开关 I 和开关 II）与 GTP 和 GDP 的磷酸及镁离子相互作用；而后两个（G4 和 G5）控制着 G 核苷酸的选择性。镁离子是水解 GTP 必需的辅因子（cofactor）。

第一个 loop 被称为磷酸结合 loop 或者 P-loop（图 5.25）。它围绕着 G 核苷酸的 α 和 β 磷酸并与之形成氢键。开关 I （G2），或者称为效应区域，以及开关 II （G3）主要通过镁离子与核苷酸的 β 和 γ 磷酸相互作用。这些 loop 的构象取决于其是否结合了 GTP 或 GDP 分子，或者说核苷酸结合位点是否是空的。效应 loop（开关 I ）参与受体的结合并且可以将构象在结合 GTP 和 GDP 的状态之间彻底地转换。开关 II 将结合核苷酸的状态传达给多结构 GTP 水解酶的其他结构域。

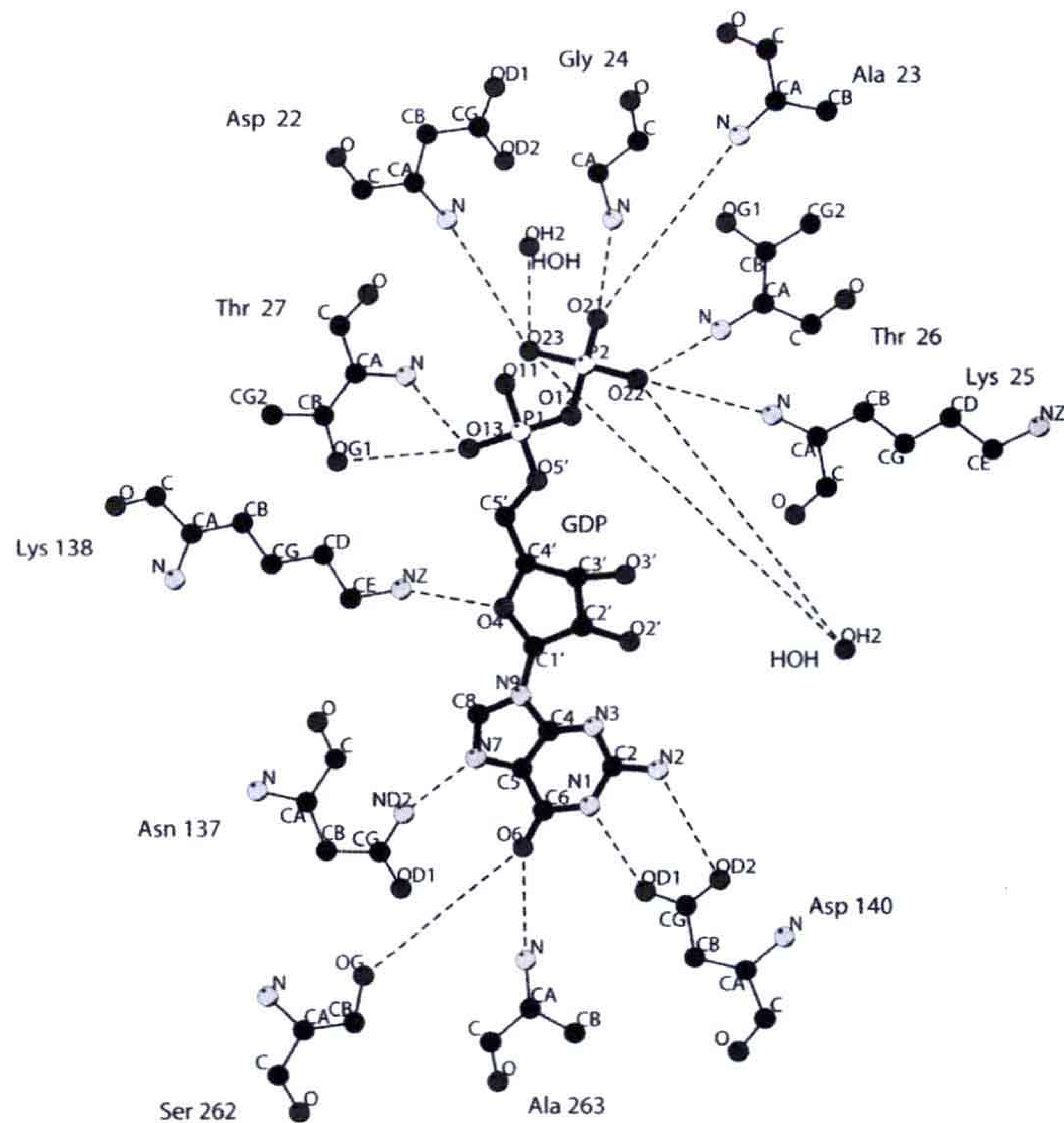


图 5.24 来自嗜热栖热菌（*T. thermophilus*）EF-G 蛋白与 GDP 的相互作用。24 位的甘氨酸、25 位的赖氨酸和 26 位的苏氨酸均属于 P-loop 中用来结合 α 和 β 磷酸的保守残基。137 位的天冬酰胺、138 位的赖氨酸和 140 位的天冬氨酸是 G4 模体的 NKXD 的一部分，用来结合鸟嘌呤。262 位丝氨酸和 263 位丙氨酸属于 G5 模体，并且也与鸟嘌呤相互作用。开关 I 和开关 II（分别为 G2 和 G3 模体）与镁离子和 γ 磷酸相互作用［感谢 Al-Karadaghi 等提供刊印授权，原文出自（1996）The structure of elongation factor G in complex with GDP: conformational flexibility and nucleotide exchange. *Structure* **4**:555-565. 版权所有：（1996）Elsevier］。

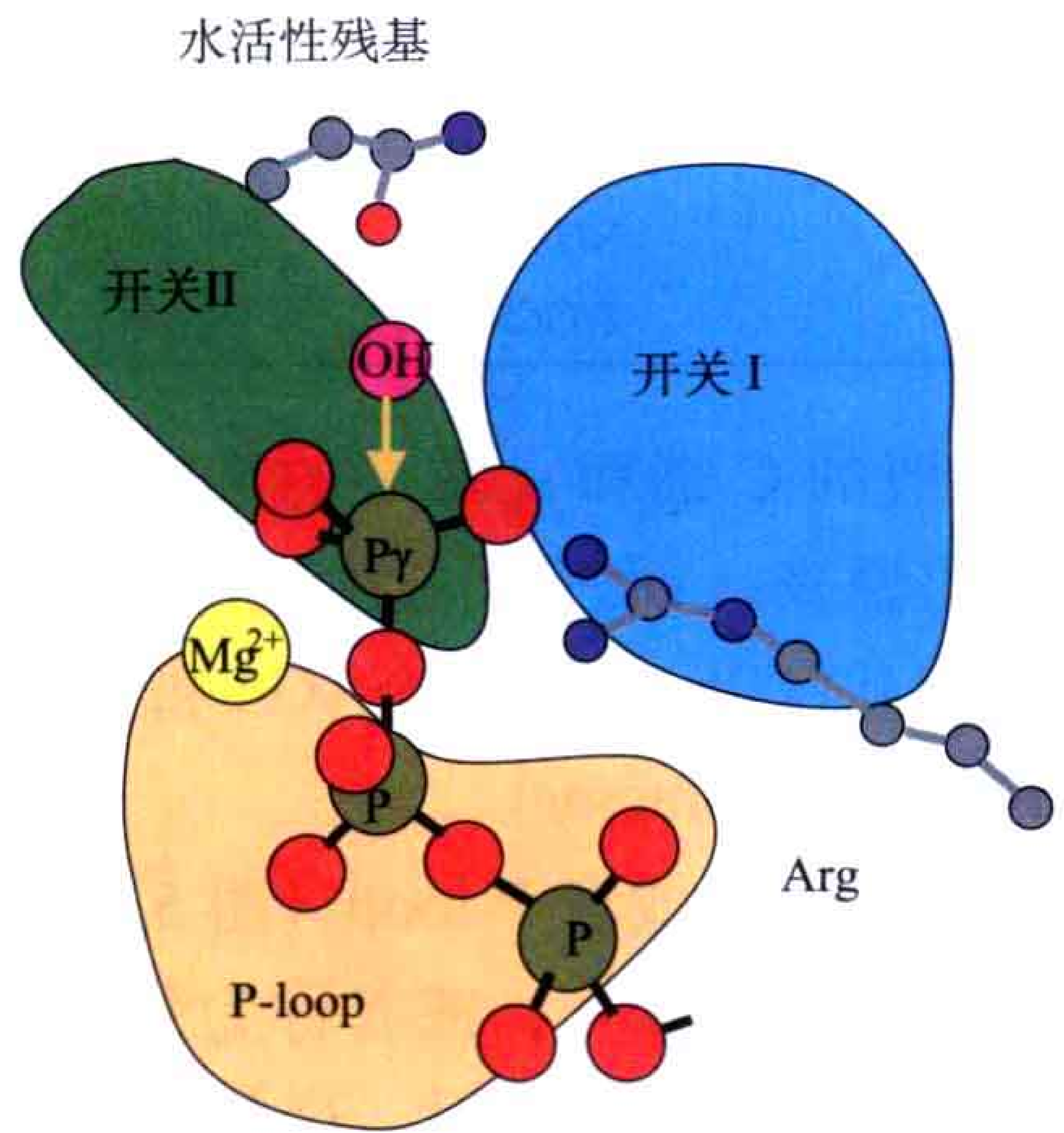


图 5.25 P-loop 与 α 和 β 磷酸的相互作用。开关 I 和开关 II 的 loop 与 γ 磷酸和镁离子相互作用且处于一个开启的状态。一个 GTP 水解酶激活蛋白（GAP）引发了 G 蛋白的一个有活性的 GTP 水解构象。这种活化一部分是由于一个与 γ 磷酸相互作用的精氨酸残基，它稳定了反应的中间状态。在这个过程中水分子被移动至靠近 γ 磷酸的位置并被去质子化，从而与磷酸发生直线的 S_N2 机理的反应。

G 蛋白可以被称为分子开关。当 G 蛋白结合 GTP 分子时处于开启的状态，它可以结合一个受体或者是效应子。这种结合可能会导致 G 蛋白与 GAP 相互作用，从而促使 G 蛋白水解其结合的 GTP 分子。当 GTP 被水解后，G 蛋白的构象转变成关闭的状态并且与效应子分离。显然，G 蛋白属于不完全的酶，它们自身的 GTP 水解酶活性很低。因此，它们必须与细胞中合适的元件相互作用从而被激活。

5.3.3.2　GTP 水解

GTP 水解的机理在许多的 G 蛋白中都已经被彻底地研究过。GAP 所扮演的角色是引发 G 蛋白的一个构象变化，使得 G 蛋白与其效应子解离并变成有活性的 GTP 水解酶。GTP 水解的催化机理在所有 GTP 水解酶中都是类似的。一个水分子，或者更具体地说是一个羟基离子，被放置于靠近 γ 磷酸且在 β 磷酸对面的位置，将被用来水解 GTP。这种水分子在 GTP 水解酶的结构中经常被发现。激活 GTP 水解酶的两个要求是稳定中间状态及去除水分子上的一个质子以激活水分子。这种去质子化使得水分子可以以直线的 S_N2 机理进攻 γ 磷酸并且将 GTP 水解成 GDP 和无机磷酸（图 5.26）。

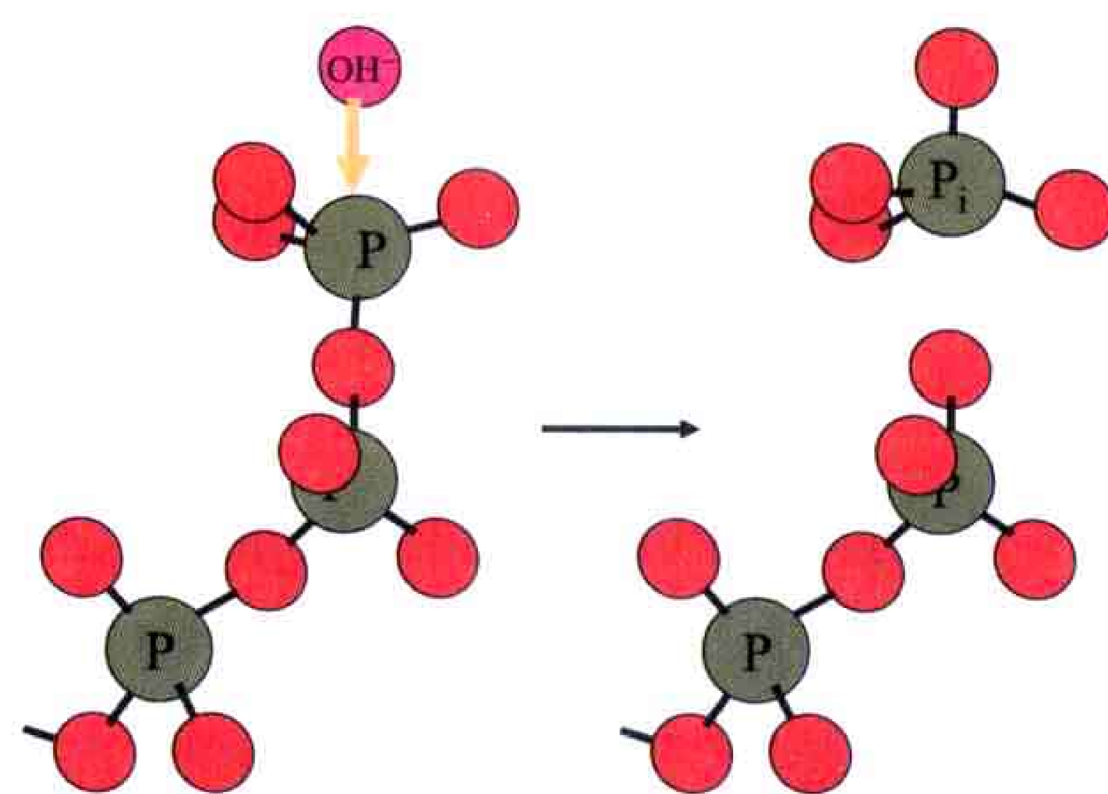


图 5.26 羟基离子进攻 ATP 或者 GTP 分子的 γ 磷酸。这个反应的结果可以想象成被强风吹得翻面的雨伞。

对于许多 G 蛋白来说有一个位于开关 II 上的谷氨酰胺残基会与水分子相互作用，并可在 GAP 的诱导下将水分子放置在适合进攻的位置。从 G 蛋白与 GDP 和氟化铝的复合体的结构中我们可以观察到 GTP 水解时不同状态的细节。通过研究活性位点附近的一些残基我们可以得出一个结论，即 γ 磷酸从水分子处吸收了一个质子，从而导致其被攻击。这种机理被称为底物诱导催化（substrate-induced catalysis），这种催化在水分子没有到位的情况下是不会发生的。G 蛋白中间状态的稳定一般是靠一个被称为精氨酸指（图 5.25）的精氨酸残基。这个精氨酸既可来自 G 蛋白本身（顺式）也可来自 GAP（反式）。

5.4　脂肪酸合成酶——一个多功能的酶

脂肪酸是生物膜中必需的分子，并且在生物系统中扮演许多其他的角色。脂肪酸的合成主要是由脂肪酸合成酶（fatty acid synthase, FAS）完成的。一些其他的酶也参与了脂肪合成的底物的生成并且还有一些酶接下来将利用合成的主产物，即棕榈酸（palmitate）以生成特定的分子用于能量储存或者一些结构用途。合成反应是通过活化的前体（precursor）的一步步延伸反应，每次加上两个碳原子。在真菌中这个合成是由

一个 $\alpha_6\beta_6$ 的分子质量为 2.6 MDa 的十二聚体执行的。而在一些生物体中这个酶则是由单独一条多肽链组成的。在脊椎动物中执行这个催化的酶是一个 α_2 的二聚体，其中每一个单体的分子质量是 270 kDa。这种酶是多结构域的蛋白质，并且具有所有脂肪酸合成中所需的酶的功能，这被称为 I 型系统(type- I system)，因此说这种酶是多功能的酶。然而在细菌和植物中，一般都分别由具有不同活性的酶来执行不同的功能，编码这些酶的基因也是独立的。这种方式就被称为 II 型 FAS 系统（FAS type- II）。细菌中的上述这些酶在功能上和结构上与 I 型 FAS 系统中不同的结构域是对应的。FAS I 型脂肪酸合成酶属于一类多功能的酶，这类酶可以重复地聚合羧酸（聚酮合成酶）或者氨基酸（非核糖体的多肽合成酶）。

5.4.1 结构

脂肪酸合成酶的核心部件是酰基载体蛋白（acyl carrier protein, ACP），它可以在不同活性位点之间搬运反应中间体。这就增加了催化位点附近的底物浓度，从而提高了催化效率。这种不同催化位点间的底物运输使得催化循环得以进行。而一旦底物的长度到达 16 ~ 18 个碳原子，就会被硫酯酶（thioesterase, TE）解离下来。目前我们得到的结构有来自细菌单独亚基的高分辨率结构、来自真菌全酶的高分辨率结构及来自猪中等分辨率的结构。

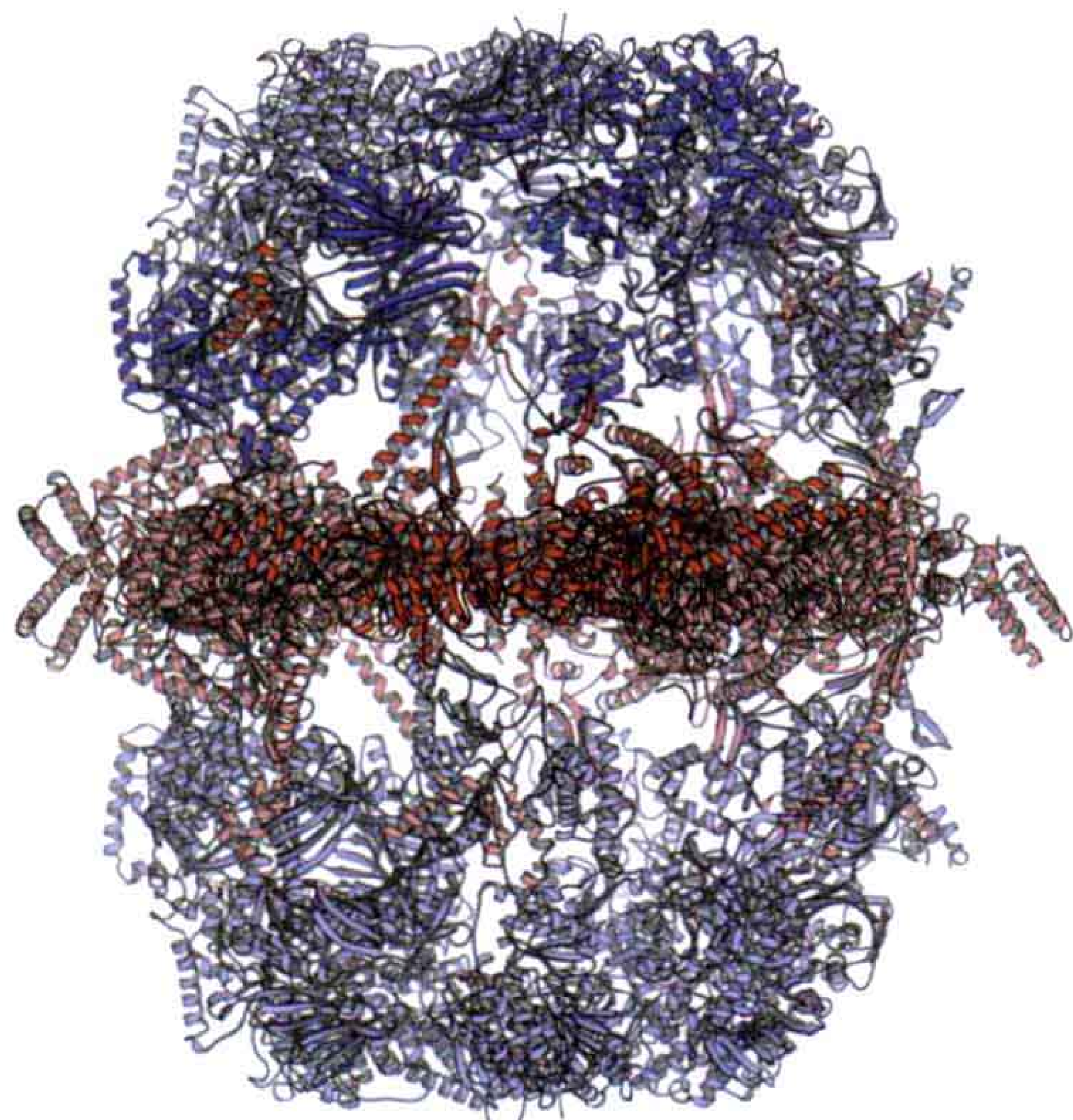


图 5.27 来自嗜热丝孢菌的脂肪酸合成酶的结构。中心的 α 亚基形成的盘状结构以红色表示，而 β 亚基形成的穹顶结构以蓝色表示。其中一个 α 亚基和一个 β 亚基以较深的颜色显示（PDB：2UV9 和 2UVA）。

从来自于真菌酿酒酵母（*S. cerevisiae*）和嗜热丝孢菌（*Thermomyces Lanuginosus*）的 FAS 结构（图 5.27）中可以看出其含有一个中心的轮状结构，包含了 6 个 α 亚基（又被称为 FAS-2，由 1878 个氨酸基组成），其上下两侧各有三个 β 亚基，围成了两个用于反应的腔室（reaction chamber）。β 亚基（又被称为 FAS-1，由 2080 个氨基酸组成）以三次轴（C_3）的对称性排列成为三聚体而 α 亚基则形成二聚体并同时也以三次轴的对称性排列（32 或 D_3 对称性，图 5.28）。6 个 α 亚基相互纠缠并且它们的 N 端分别与中心环上下两侧的穹顶形成连接（三个与上侧而另三个与下侧），而每个穹顶结构由三个 β 亚基构成。整个结构高 270Å，宽 250Å。反应腔室含有一定数量的“窗户”与外界相连。并且两个反应腔室之间也有“窗户”连接。

来自于猪的脂肪酸合成酶是由两个相同的 270 kDa 的链构成的，而它的结构大不相同。其总体的形状是一个 X 形，含有两条腿和两只手臂。尺寸是 210Å × 180Å × 90Å

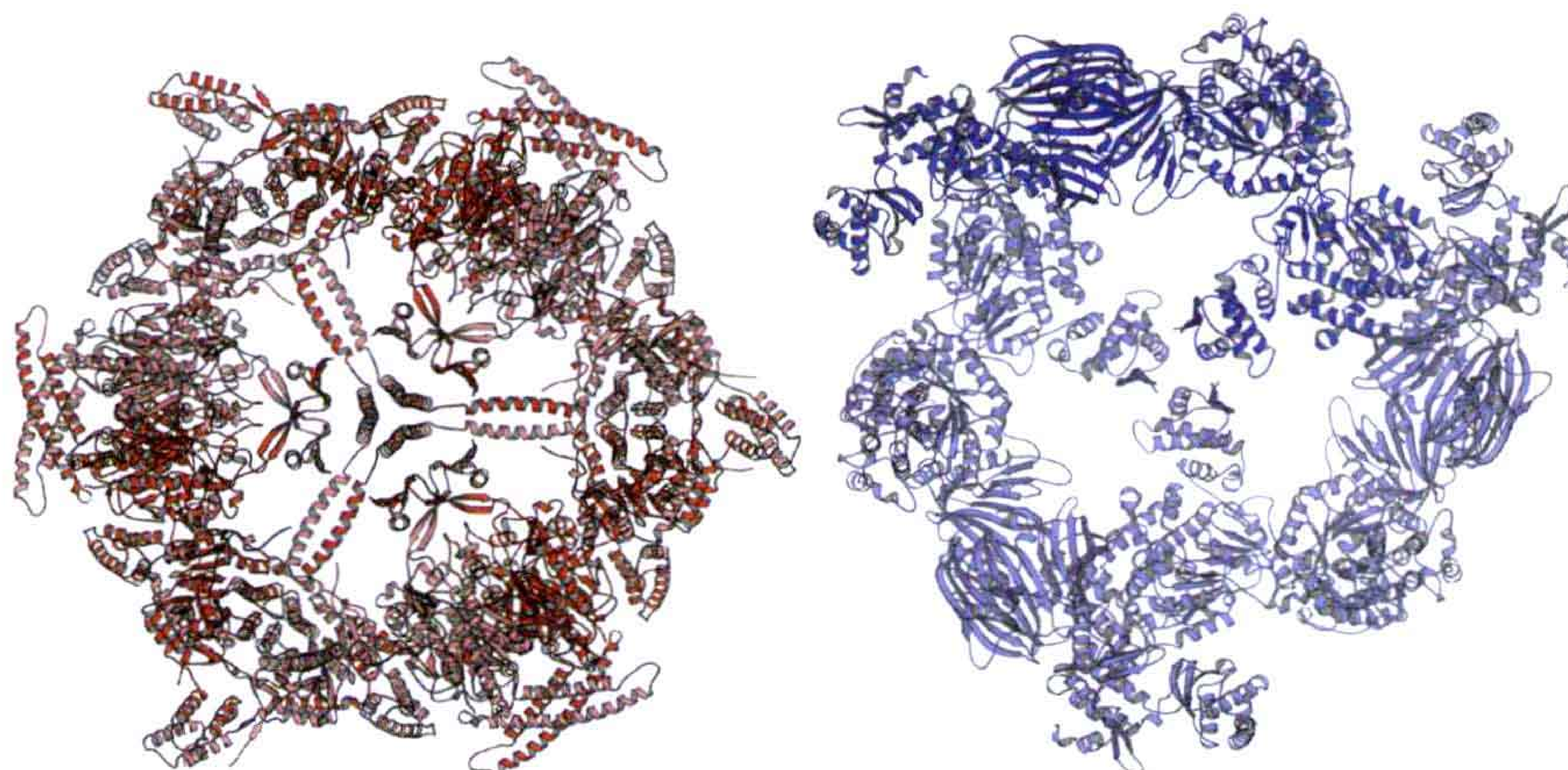


图 5.28　左：中央的盘状结构由 6 个相互纠缠的 α 亚基组成并以 D_3 对称性排列；右：两个穹顶中的每一个都由三个 β 亚基构成。两图的视角均是顺着三次轴方向的，并且与图 5.27 的视角垂直。三个 α 亚基和一个 β 亚基被高亮显示。

且有一个垂直的二重轴。鉴于这个结构的分辨率相对较低，我们将不深入讨论这一型的脂肪酸合成酶。

5.4.2　酶学功能

5.4.2.1　功能结构域

酵母的脂肪酸合成酶有 8 个功能结构域：α 亚基的 ACP（酰基载体蛋白）、KR（ketoacyl reductase，酮脂酰还原酶）、KS（ketoacyl synthase，酮脂酰合成酶）、PT（phosphopantheine trasferase，磷酸泛酰巯基乙胺转移酶）及 β 亚基的 AT（acetyl transferase，乙酰转移酶）、ER（enoyl reductase，烯酰还原酶）、DH（dehydratase，脱水酶）和 MPT（malonyl/palmitoyl transferase，丙二酰 / 棕榈酰转移酶）。其中 MPT 结构域既有来自于 β 亚基 C 端的部件，也有来自于 α 亚基 N 端的部件（图 5.29）。另外，β 亚基还含有 4 个没有催化活性的结构域。

5.4.2.2　催化

图 5.30 所示为脂肪酸合成酶中每个结构域在合成脂肪酸时所扮演的角色。β 亚基的 AT 和 MPT 结构域的折叠方式类似于铁氧还蛋白。它们的活性位点包含了组氨酸和丝氨酸残基，均位于一个裂隙中，但由于功能和底物的不同而有所差异。特别是 MPT 在其催化口袋的底部有一个保守的精氨酸残基，用于与丙二酰基的羧基基团相互作用。AT 的催化裂隙比较狭窄，阻止了长链底物的结合，而 MPT 的催化裂隙较大且疏水性更强，从而允许较大底物的结合。

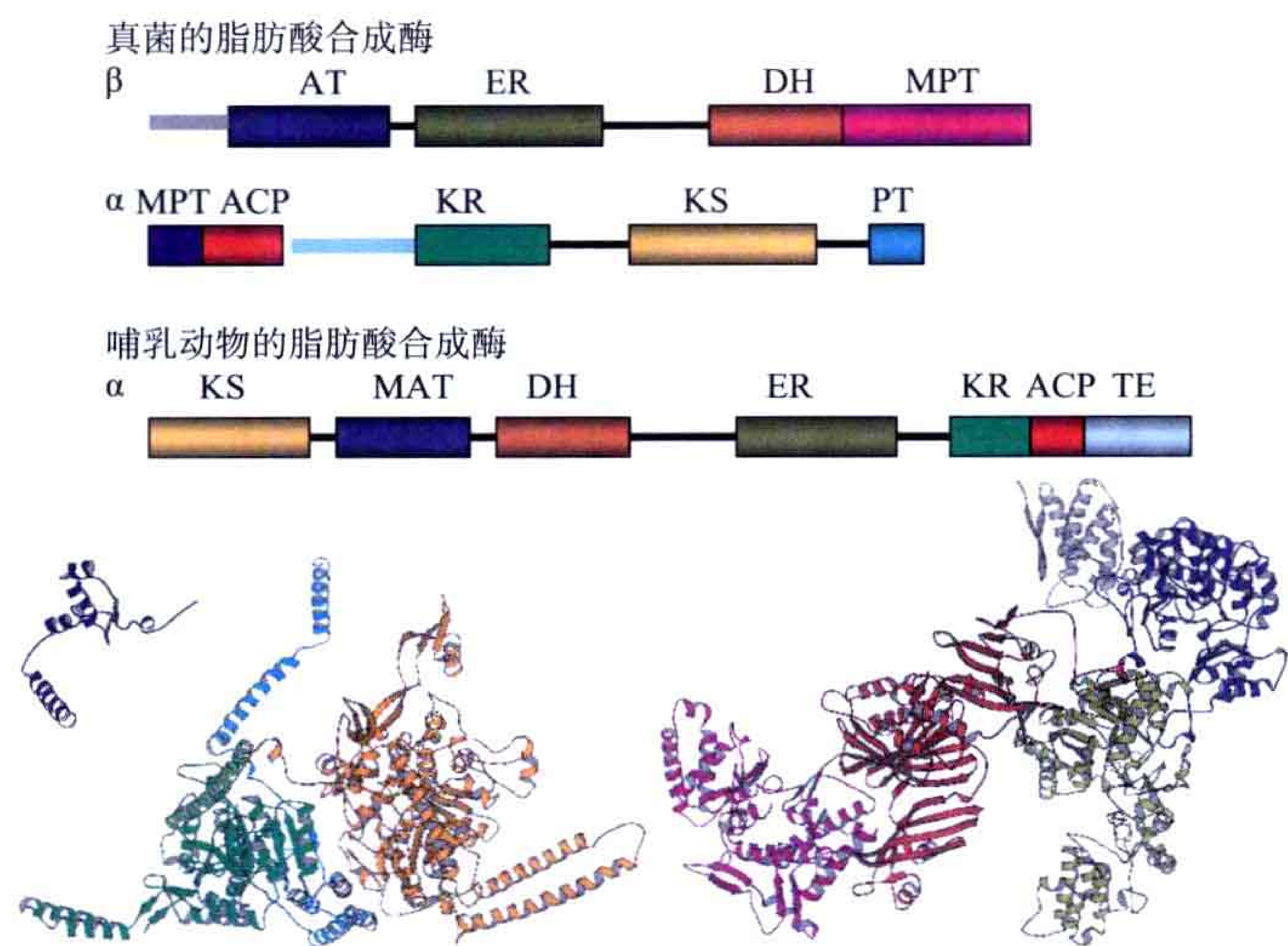


图 5.29 上：真核生物的脂肪酸合成酶的结构域组成；下：真菌的脂肪酸合成酶由 6 个 α 亚基（下图左）和 6 个 β 亚基（下图右）组成，而哺乳动物的脂肪酸合成酶仅含有两个相同的亚基。尽管两个形式的脂肪酸合成酶在演化上是相关的，但其功能结构域的排列顺序是不一样的。在真菌的脂肪酸合成酶中，丙二酰 / 棕榈酰转移酶是由两个亚基共同组成的。而在哺乳动物中，MAT 结构域可以同时执行乙酰丙二酰的转移功能。

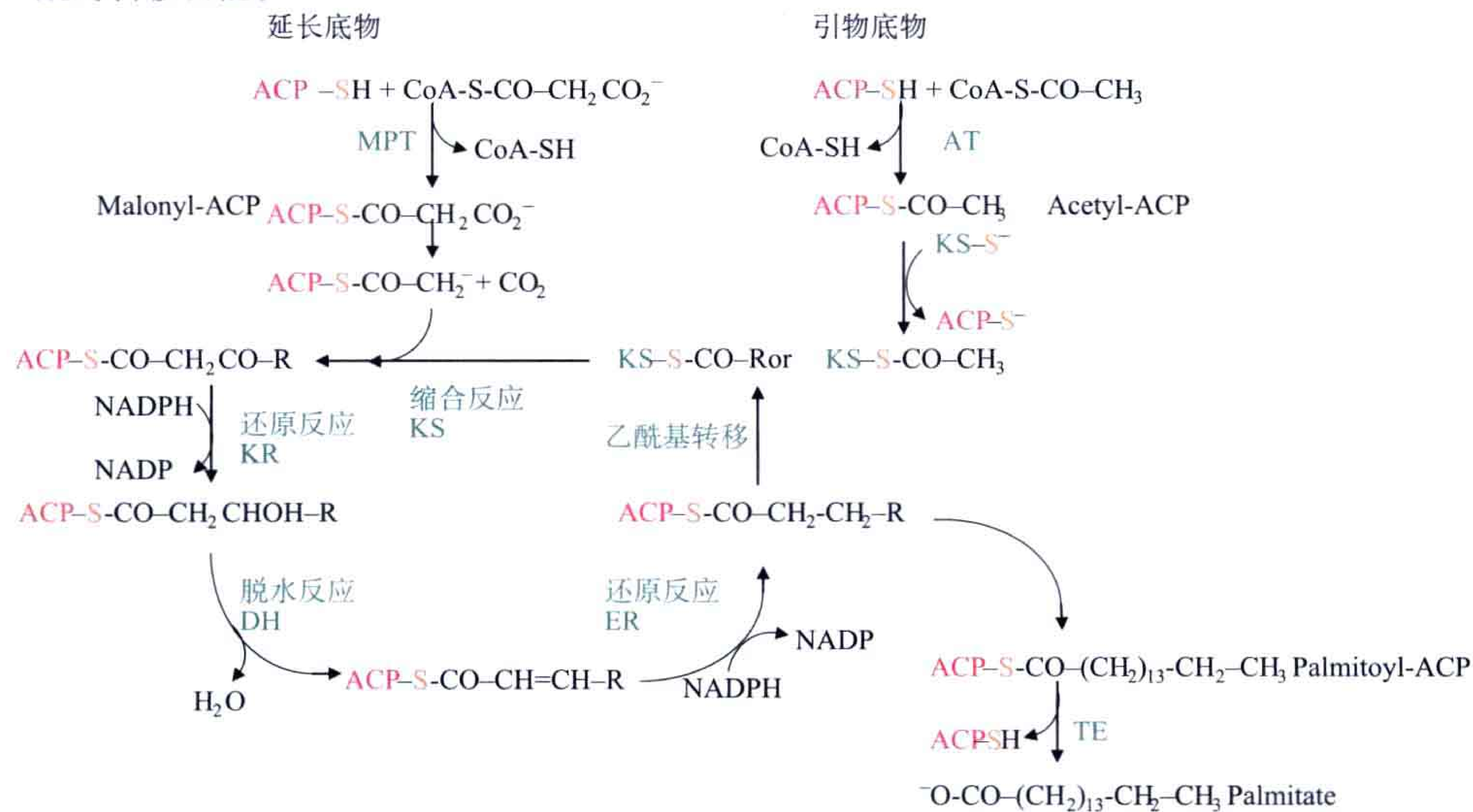


图 5.30 合成脂肪酸的主要步骤及一些催化结构域（绿色）在其中扮演的角色。合成的最初步骤是将一个乙酰基从辅酶 A（coenzyme A, CoA）上转移至 ACP（红色）。用于连接的基团是一个长 18Å 的磷酸泛酰巯基乙胺基团，其黏附在 ACP 的丝氨酸上。乙酰基底物通过与连接基团末端的 S 原子形成硫脂键连接 ACP。乙酰基团是脂肪酸合成最初的两碳底物并随后被转移到酮脂酰合成酶（KS）活性位点的半胱氨酸处。然后 ACP 连接了一个连电的丙二酰基团。在脱羧后便产生了一个活性乙酰碳负离子，从而可以与黏附在 KS 的半胱氨酸上的乙酰基或者脂肪酸反应。随后底物分子被转移至酮脂酰还原酶（KR），以 NADPH 作为还原剂发生还原反应。脱水酶（DH）将底物分子脱去一分子水，接下来烯酰还原酶用另一分子的 NADPH 将其还原。此时底物分子比之前长了两个碳原子并且又被连接在 KS 结构域上。这个反应过程会一直进行直到到达 16 或 18 个碳原子的长度，此时脂肪酸会被 MPT 转回到辅酶 A 上。在哺乳动物中还会有一个酯转移酶结构域（tarsesterase domain, TE）通过水解反应将底物最终释放。

KS 催化乙酰基和丙二酰基的缩合反应。KS 结构域是 α 亚基的一部分且在中心轮状结构中以二聚体的形式出现，但其催化位点分别指向两个反应腔室。它的折叠方式与硫解酶是一致的。KS 催化所有的缩合步骤，因此结合在其活性位点半胱氨酸上并为反应提供乙酰基的底物既可以是短链的也可以是长链的（C_2 到 C_{16}）。

KR 结构域以 NADPH 为还原剂还原 β 碳，它的折叠方式为 Rossmann 折叠。DH 结构域负责从羟基乙酰基 -ACP 上去除一个水分子。它有一个长的疏水孔道，从而允许长链底物的结合。ER 结构域也是一个二聚体，它催化整个延伸循环中最终的还原反应步骤，使用的还原剂同样是 NADPH。它的结构中含有一个 TIM 桶（TIM-barrel）折叠。这个结构域中还含有一个黄素单核苷酸（flavin mononucleotide, FMN）结合在 TIM 桶的底部。

ACP 结构域负责将底物分子在反应腔室中的不同活性位点之间进行运输。它是 α 亚基的一部分并且由两段柔性的连接肽段（linker）形成铰链状的结构（图 5.31）。由于其较强的柔性，得到 ACP 具体的结构是一件很困难的事情。然而，在酿酒酵母的脂肪酸合成酶的结构中它被固定在了 KS 的活性位点处。这个 ACP 分子质量约为 18 kDa 且完全是由螺旋结构组成。两个连接肽段与 α 亚基组成的盘状结构的中心相接，而外围锚点连接在 α 亚基的 N 端。长为 18 kDa 的磷酸泛酰巯基乙胺（PPT）是与底物共价结构的化学基团。它与 ACP 的丝氨酸残基连接，而负责完成这个连接反应的是 PT 结构域。PPT 基团位于 ACP 上两个连接钛段对面的位置，这个安排使得 PPT 基团可以伸展到最大的距离。

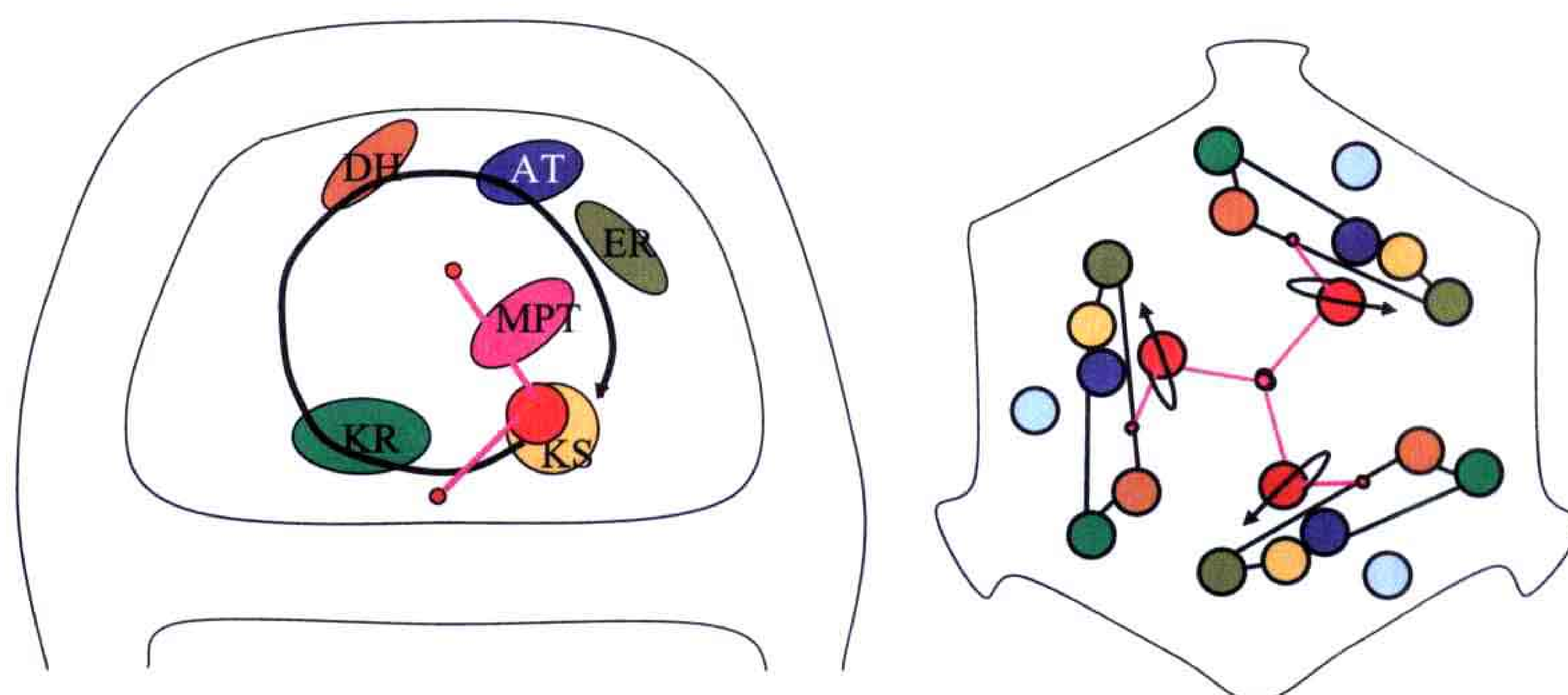


图 5.31 FAS 的反应腔室中各活性位点的分布。左：以横截面显示的反应腔室。ACP 及其两个连接肽段以红色显示。其中较下方的连接肽段与中心盘状结构连接而较上方的（又称外围的）连接肽段连接在 α 亚基的 N 端。黑色箭头显示的是底物在反应循环中经过 4 个活性位点的走向。右：反应腔室的俯视视角，显示了全部的三个 ACP 及与之相应的活性位点。

6 个不同催化结构域（AT、MPT、KS、KR、DH 和 ER）的活性位点全部都朝向两个反应腔室的内侧，而 PT 结构域则位于反应腔室的外侧。每一个腔室包含三个 ACP 结构域。ACP 的位置及反应腔室中各活性位点的排布方式使得脂肪酸的催化十分有效率。ACP 上柔性的连接肽段使得所有的活性位点都可以十分便捷地得到相应的底物。当底物被 AT 或 MPT 装载到 ACP 上后，ACP 结构域将以顺时针顺序（从反应腔室内看）依序经过相应的活性位点。经过这样一个循环后底物的乙酰链被加长了两个碳原子并

且被送回了 KS 的半胱氨酸处，在此 ACP 会从 MPT 结构域处结合一个新的丙二酰。由于有两个连接肽段，ACP 结构域更倾向于沿着圆周的方向移动，这个移动路径正好与各活性位点的排布位置一致（图 5.31）。每一个 ACP 结构域与来自两个 α 亚基和两个 β 亚基的活性位点相互作用。

延伸阅读（5.1 节）

综述文章

Krishnamurthy VM, Kaufman GK, Urbach AR, Gitlin I, Gudiken KL, Weibel DB, Whitesides GM. (2008) Carbonic anhydrase as a model for biophysical and protein-ligand binding. *Chem Rev* **108**: 946-1051.

Lindskog S, Liljas A. (1993) Carbonic anhydrase and the role of orientation in catalysis. *Curr Op Struct Biol* **3**: 915-920.

Liljas A, Håkansson K, Jonsson BH, Xue Y. (1994) Inhibition and catalysis of carbonic anhydrase. Recent crystallographic analyses. *Eur J Biochem* **219**: 1-10.

Liljas A, Laurberg M. (2000) A wheel discovered three times. The molecular structure of the three carbonic anhydrases. *EMBO Rep* **1**: 16-17.

Tripp BC, Smith K, Ferry JG. (2001) Carbonic anhydrase: New insight for an ancient enzyme. *J Biol Chem* **276**: 48615-48618.

延伸阅读（5.2 节）

原始文献

Xu H, Faber C, Uchiki T, Fairman JW, Racka J, Dealwis C. (2006) Structures of eukaryotic ribonucleotide reductase I provide insight into dNTP regulation. *Proc Natl Acad Sci USA* **103**: 4022-4027.

综述文章

Eklund H, Uhlin U, Farnegardh M, Logan DT, Nordlund P. (2001) Structure and function

of the radical enzyme ribonucleotide reductase. *Prog Biophys Mol Biol* **77**: 177-268.

Jordan A, Reichard P. (1998) Ribonucleotide reductases. *Annu Rev Biochem* **67**: 71-98.
Larsson KM, Jordan A, Eliasson R, Reichard P, Logan DT, Nordlund P. (2004) Structural mechanism of allosteric substrate specificity regulation in a ribonucleotide reductase. *Nat Struct Mol Biol* **11**: 1142-1149.

Nordlund P, Reichard P. (2006) Ribonucleotide reductases. *Annu Rev Biochem* **75**: 681-706.

延伸阅读（5.3 节）

原始文献

Ariga T, Muneyuki E, Yoshida M. (2007) F1-ATPase rotates by an asymmetric, sequential mechanism using all three catalytic subunits. *Natl Struct Mol Biol* **14**: 841-846.

Abrahams JP, Leslie AGW, Lutter R, Walker JE. (1994) Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. *Nature* **370**: 621-628.

Bourne HR, Sanders DA, McCormick F. (1991) The GTPase superfamily: Conserved structure and molecular mechanism. *Nature* **349**: 117-127.

deVos A, Tong L, Milburn MV, Mathias PM, Jancarik J, Noguchi S, Nishimura S, Miura K, Ohtsuka E, Kim S-H. (1988) Three-dimensional structure of an oncogene protein: Catalytic domain of human c-H-ras p21. Science 232: 1127-1132. Noji H, Yasuda R, Yoshida M, Kinosita K, Jr. (1997) Direct observation of the rotation of F1-ATPase. *Nature* **386**: 299-302.

Pai EF, Kabsch W, Krengel U, Holmes KC, Johns J, Wittinghofer A. (1989). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35Å resolution: implications for the mechanism of GTP hydrolysis. *Nature* **341**: 209-214.

Kabaleeswaran V, Puri N, Walker JE, Leslie ASW, Mueller DM. (2006) Novel features of the rotatory catalytic mechanism revealed in the structure of yeast F_1 ATPase. *EMBO J* **25**: 5433-5442.

Pan X, Eathiraj S, Munson M, Lambright DG. (2006) TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature* **442**: 303-306.

综述文章

Dimroth P, von Ballamos C, Meier T. (2006) Catalytical and mechanical cycles in F-ATP synthases. *EMBO Rep* **7**: 276-282.

Gao YQ, Yang W, Karplus M. (2006) Astructure-based model for the synthesis and hydrolysis of ATP by F_1-ATPase. *Cell* **123**: 195-205.

Weber J. (2007) ATP synthase—The structure of the stator stalk. *Trends Biochem Sci* **32**: 53-56.

Tucker PA, Sallai L. (2007) The AAA+ superfamily—A myriad of motions. *Curr Op Struct Biol* **17**: 641-652.

延伸阅读（5.4 节）

原始文献

Jenni S, Leibundgut M, Boehringer D, Frick C, Mikolásek B, Ban N. (2007) Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**: 254-261.

Leibundgut M, Jenni S, Frick C, Ban N. (2007) Structural basis for substrate delivery by acyl carrier protein in yeast fatty acid synthase. *Science* **316**: 288-290.

Lomakin IB, Xiong Y, Steitz TA. (2007) The crystal structure of fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **129**: 319-332.

Maier T, Jenni S, Ban N. (2006) Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**: 1258-1262.

（曹　骎　译　刘翔宇　苏晓东　校）

第六章

DNA 代谢——复制与重组

6.1 DNA 复制

DNA 的代谢是细胞最重要的活动之一，它包括复制、修复、重组和降解。成百上千的酶参与了 DNA 的代谢，从而将遗传信息代代相传，在此我们将主要讲解其中的一部分酶。

马休 · 梅塞尔伸（Matthew Meselson）和富兰克林·斯塔尔（Flanklin Stahl）认为 DNA 双螺旋中两条平行的链在 DNA 复制中都作为合成新链的模板，生成的子代 DNA 双链中一条来自亲代 DNA，另一条是合成的新链，这被称为半保留复制（semi-conservative replication）。除模板 DNA 外，DNA 的复制还需要底物核酸前体（NTP）和 DNA 聚合酶（DNA polymerase）的参与。

复制过程分为两步：① DNA 双链解旋和分离，使其可以与新碱基配对；②新的核苷酸通过与模板碱基互补配对共价结合在新链末端。新的核苷三磷酸的磷酸连接在脱氧核糖的 5′ 端，加在新链 DNA 末端脱氧核糖的游离羟基上。解旋酶（helicase）和 DNA 聚合酶分别参与这两步反应，而这两个酶还需要有其他几个蛋白质辅助完成其功能。

在 DNA 复制过程中，DNA 聚合酶沿着模板链的 3′ 端向 5′ 端催化 DNA 合成反应，而新合成的 DNA 链由 5′ 端起始，向 3′ 端延伸。DNA 双链在复制过程中需要解离成两条模板链，在 DNA 上形成一个分支点，这被称为复制叉（replication fork，图 6.1）。其中一条链从 3′ 端到 5′ 端的方向，这条链作为引导链使得复制叉向前推进；另一条链是从 5′ 端到 3′ 端，被称为滞后链，这条链也是沿着 3′ 端到 5′ 端的方向被复制。复制是以许多片段的形式完成的，每个片段是从复制叉起始，到上一个复制片段的 3′ 端结束。这些片段被称为冈崎片段，可以通过 DNA 连接酶将它们连接在一起。

在细胞中，每个单独的 DNA 分子（如一条染色体），其序列中都含有至少一个复制起点（the origin of replication）作为 DNA 复制的起始位点。在大肠杆菌的染色体中，通常只存在一个复制起点；而在真核生物的线性染色体中可能存在很多复制起点，并可以同时启用而形成很多个复制叉。

复制复合体（replication complex）是一个巨大的多蛋白聚体，它参与催化 DNA 的复制。复制复合体结合在复制起点，并打开复制起点区 DNA 的双链结构。然后，DNA

复制从两个方向同时启动，形成两个移动的复制叉。新链是以母链为模板通过互补配对合成的。

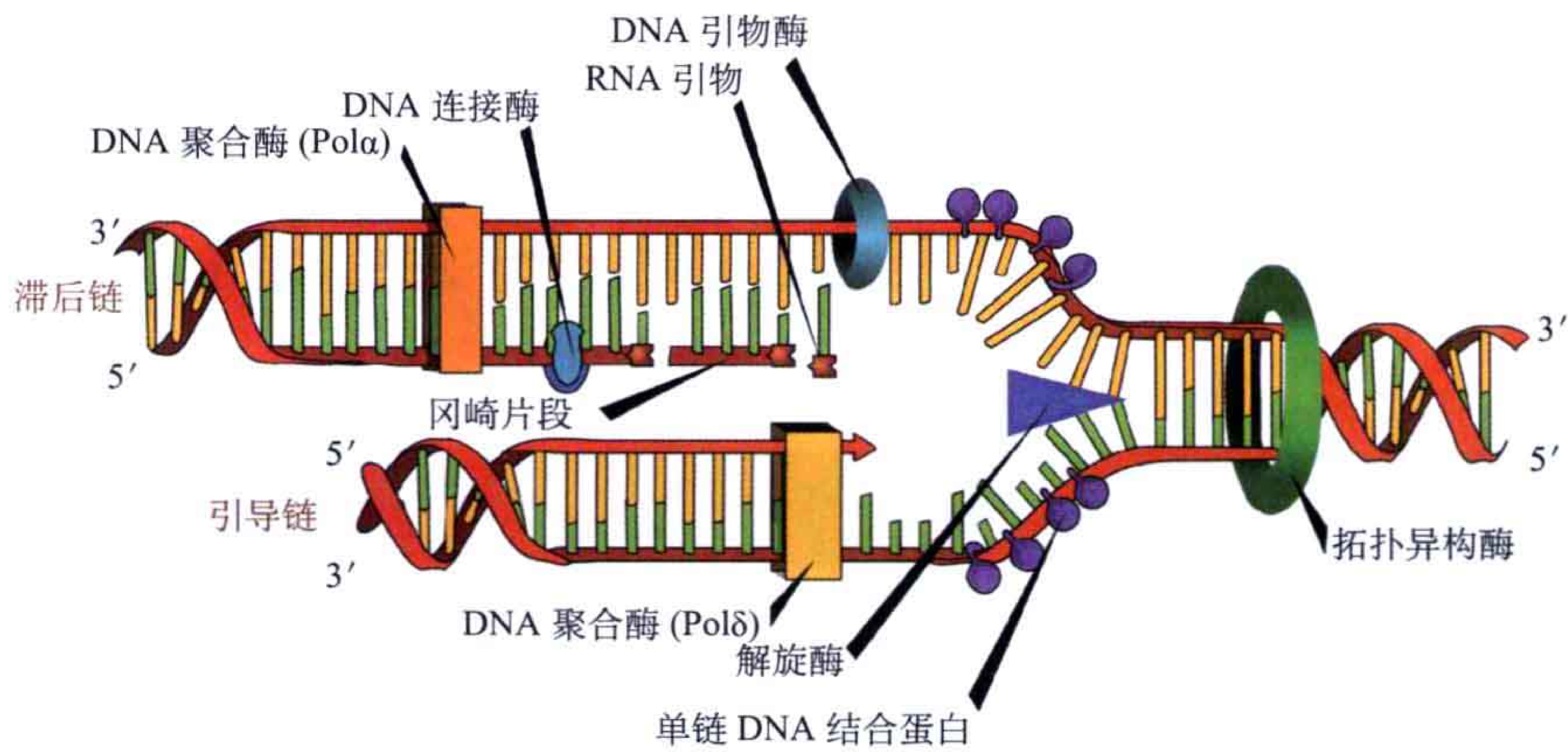


图 6.1 真核生物中参与 DNA 复制的元件。该图来自于 Wikipedia，由 Mariana Ruiz Villareal 提供。

6.1.1 DNA 解旋酶

在识别 DNA 复制起点时，DNA 解旋酶通过 ATP 水解释放的能量破坏 DNA 双链结合的作用力，在此过程中，单链 DNA 结合蛋白帮助 DNA 解旋酶来稳定打开的单链 DNA。另外，DNA 拓扑异构酶协助复制叉的移动，释放螺旋张力。DNA 拓扑异构酶切断 DNA 的一条链，使其可以绕另一条链旋转。解旋酶是分子马达，它偶联三磷酸腺苷（ATP）水解的能量用于解开和改变 DNA 双螺旋。解旋酶不仅在 DNA 的复制中必不可少，而且对于重组、修复、转录和翻译都是不可或缺的。根据解旋酶模体保守程度和其在核酸模板上的移动方向，解旋酶可以被分为几个不同的家族（表 6.1）。

表 6.1 解旋酶家族

家族	*E. coli* 中的酶	功能	此家族中其他解旋酶
超家族 1	UvrD, Rep	DNA 修复、复制	PcrA（细菌重组）
超家族 2	RecQ, UvrB	DNA 修复	eIF4 翻译因子
超家族 3	—	DNA 复制	动物病毒
DnaB 家族	DnaB	DNA 复制中的解旋酶六聚体	噬菌体复制
Rho 样家族	Rho	转录终止	

RecQ 家族的解旋酶存在于所有的物种中，主要参与 DNA 的修复。这种解旋酶有三个保守的结构域：解旋酶结构域、RecQ-C 端（RecQ-C-terminal，RecQ-Ct）结构域，以及解旋酶和 RNA 酶样 C 端（helicase-and-RNaseD-like-C-terminal，HRDC）结构域。相似的解旋酶结构域也存在于解旋酶超家族 1 中，这种结构域由 7 个保守的模块构成。这些模块可以利用核苷酸水解的能量来解开两条互补的 DNA 和（或）RNA 链。这些

模体中的一部分，包括 P-loop 模体 GXXXXGKS，提供一个 ATP 和其他聚核苷酸的结合位点（见 5.3 节）。解旋酶结构域实际上是由两个十分相似的结构域组成的，而这其中的每个结构域都和细菌重组蛋白 RecA 有结构相似性（图 6.2），这种 RecA 样模块是解旋酶马达结构域的核心。RecQ-Ct 和 HRDC 结构域参与控制底物特异性及与其他复制复合体蛋白的相互作用。目前，如何将 ATP 水解的能量转化用于 DNA 解链尚不清楚，但是很有可能是两个 RecA 样结构域之间相对趋向的改变参与了这一过程。

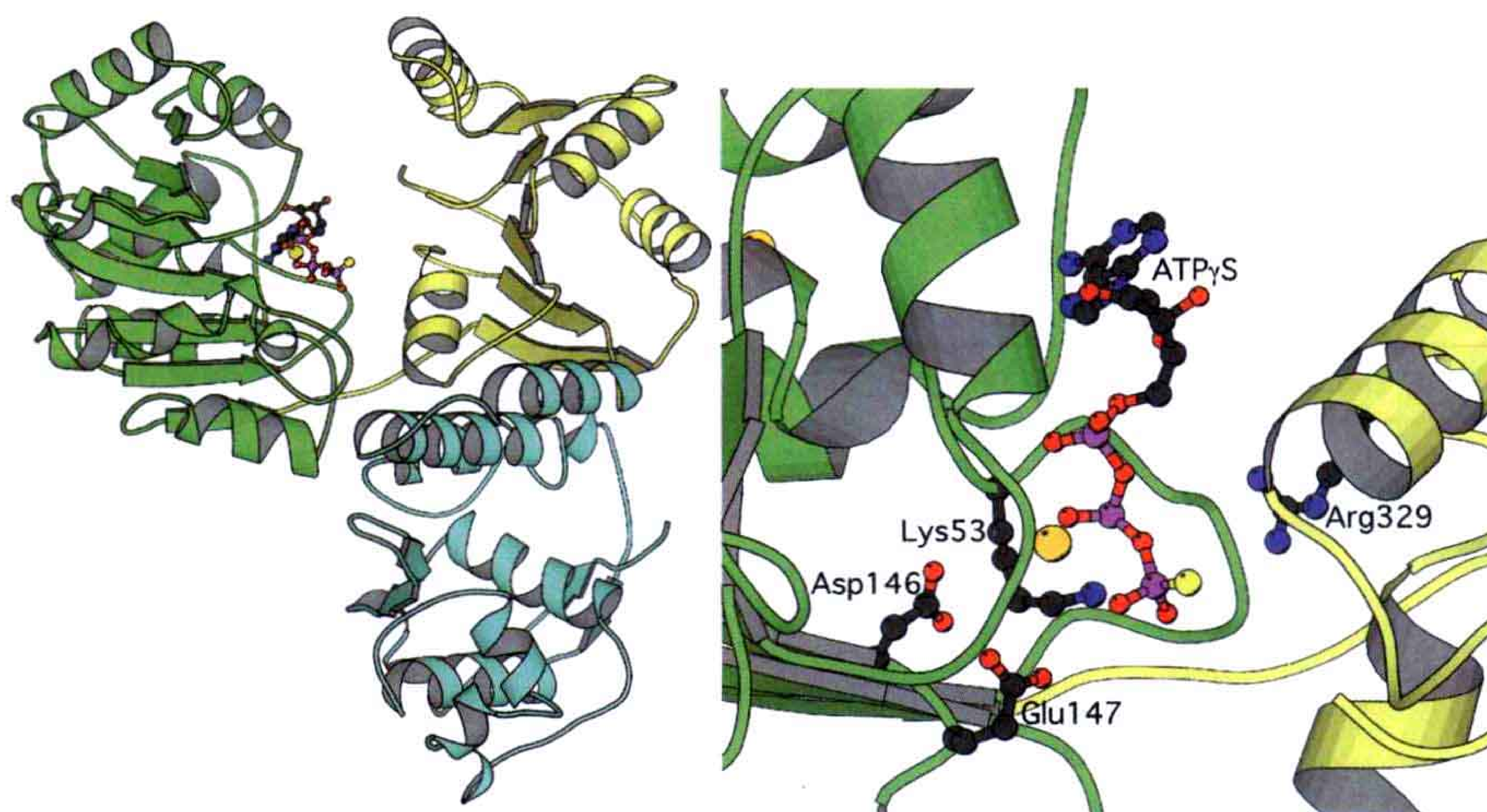


图 6.2 左：大肠杆菌 RecQ 的结构。RecA 样解旋酶结构域显示为绿色和黄色，RecQ-Ct 结构域显示为青绿色。右：ATP 结合位点及参与催化的残基。Lys53 是来自于 P-loop，Asp146 和 Glu147 来自第二个保守模体（DEXH）。与其他含 P-loop 的蛋白质的“精氨酸指”（arginine finger）（PDB：1OYY）类似，另一个结构域的 Arg329 靠近活性位点。

细菌中主要参与 DNA 复制的解旋酶属于 DnaB 家族，这个家族的解旋酶为六聚体，具有一个单一的 C 端 RecA 样模块和一个 N 端部分。

6.1.2 DNA 聚合酶

6.1.2.1 DNA 聚合酶的功能与分类

物种的稳定性取决于精确的复制及如何保存其遗传物质，同时允许存在较低的突变率。一般来说，在复制过程中每 10^9~10^{10} 的碱基中会产生一个错误。这种极高的准确度是通过精确的核苷酸插入和去除插入错误的核苷酸来保持的。DNA 聚合酶是在 DNA 复制、修复和重组中保证准确度的关键酶。总体来看，DNA 聚合酶像是一个张开的右手，由手掌、拇指和四指组成。除了 DNA 聚合的活性，很多 DNA 聚合酶还具有核酸外切酶的活性（5′ → 3′ 和 3′ → 5′）。

自 1956 年科恩伯格（Kornberg）及其同事发现了第一个 DNA 聚合酶后，又有很多 DNA 聚合酶被发现。自然界似乎通过利用不同的 DNA 聚合酶行使相似的功能创造了一个安全稳定的 DNA 复制机制。例如，在真核生物中正常的 DNA 复制需要三个 DNA 聚合酶 Polα、β、ε 的参与，而其他的 DNA 聚合酶则参与 DNA 的修复和其他需要处理受损 DNA 的机制中。基于一级序列的同源性和晶体结构分析，DNA 聚合酶可以被分为 7 类：A、B、C、D、X、Y 和 RT（表 6.2）。真核生物的 DNA 聚合酶 α、β、ε 属于 B 类。与其他 6 种以 DNA 为模板的聚合酶不同，RT 类 DNA 聚合酶可以将单链的病毒 RNA 基因组转化成双链的 DNA 插入宿主的基因组中（原病毒 provirus）。端粒酶（telomerase，见 6.1.3 节）也属于 RT 家族。一般来说，A、B、X、Y 和 RT 类 DNA 聚合酶有着相同的复制机制。

表 6.2　细菌和哺乳动物中的 DNA 聚合酶示例

类别	*E. coli* 中的酶	哺乳动物中的酶	其他酶	主要功能
A	Pol Ⅰ			DNA 修复
		Pol γ		线粒体 DNA 复制
			T7 聚合酶	噬菌体 DNA 复制
B	Pol Ⅱ			损伤 DNA 的复制
		Pol α		引物酶
		Pol δ		主要的复制酶（滞后链）
		Pol ε		主要的复制酶（引导链）
			RB69 聚合酶	噬菌体 DNA 复制（引导链和滞后链）
C	Pol Ⅲ			15 个亚基。引导链和滞后链的复制
D				古细菌的复制
X		Pol β		损伤 DNA 的修复
Y	DinB			旁路合成
			Pol η	旁路合成
RT			反转录酶	RNA 依赖的 DNA 合成

注：哺乳动物中还有几个其他的 DNA 聚合酶。此分类是基于序列的相似性。C 类 DNA 聚合酶在细菌中是主要的复制酶，但不存在于真核生物中。

DNA 聚合酶不能自行起始合成 DNA 新链，它需要一个辅助 DNA 或 RNA，其也被称为引物（primer）。通常，引物是一个由 RNA 聚合酶或被称为引物酶（primase）预先合成的短 RNA 链。DNA 聚合酶将核苷酸加在引物上直至复制完成。然后 RNA 引物需要被降解并被 DNA 片段取代，从而使复制最终合成的新链中仅有 DNA。

第一个被鉴定的 DNA 聚合酶，大肠杆菌的 DNA 聚合酶 Ⅰ，显示了 DNA 聚合酶的基本结构。该酶有 109 kDa，含有几个参与 DNA 合成反应的结构域，并同时具有 5′ → 3′ 和 3′ → 5′ 外切核酸酶活性。其 C 端缺少 5′ → 3′ 外切酶活性的部分，称为 Klenow 片段，被广泛地应用于实验室分子生物学。Klenow 片段是第一个结构，被

解析的 DNA 聚合酶。这种基本类似右手的结构，后来被发现存在于几乎所有的 DNA 聚合酶中（图 6.3）。直接参与 DNA 聚合反应的活性位点，位于手掌亚结构域，构成了缺口的底座，被拇指及四指亚结构域环绕。构成拇指的部分参与 DNA 结合，而四指则参与核苷酸的识别和结合。

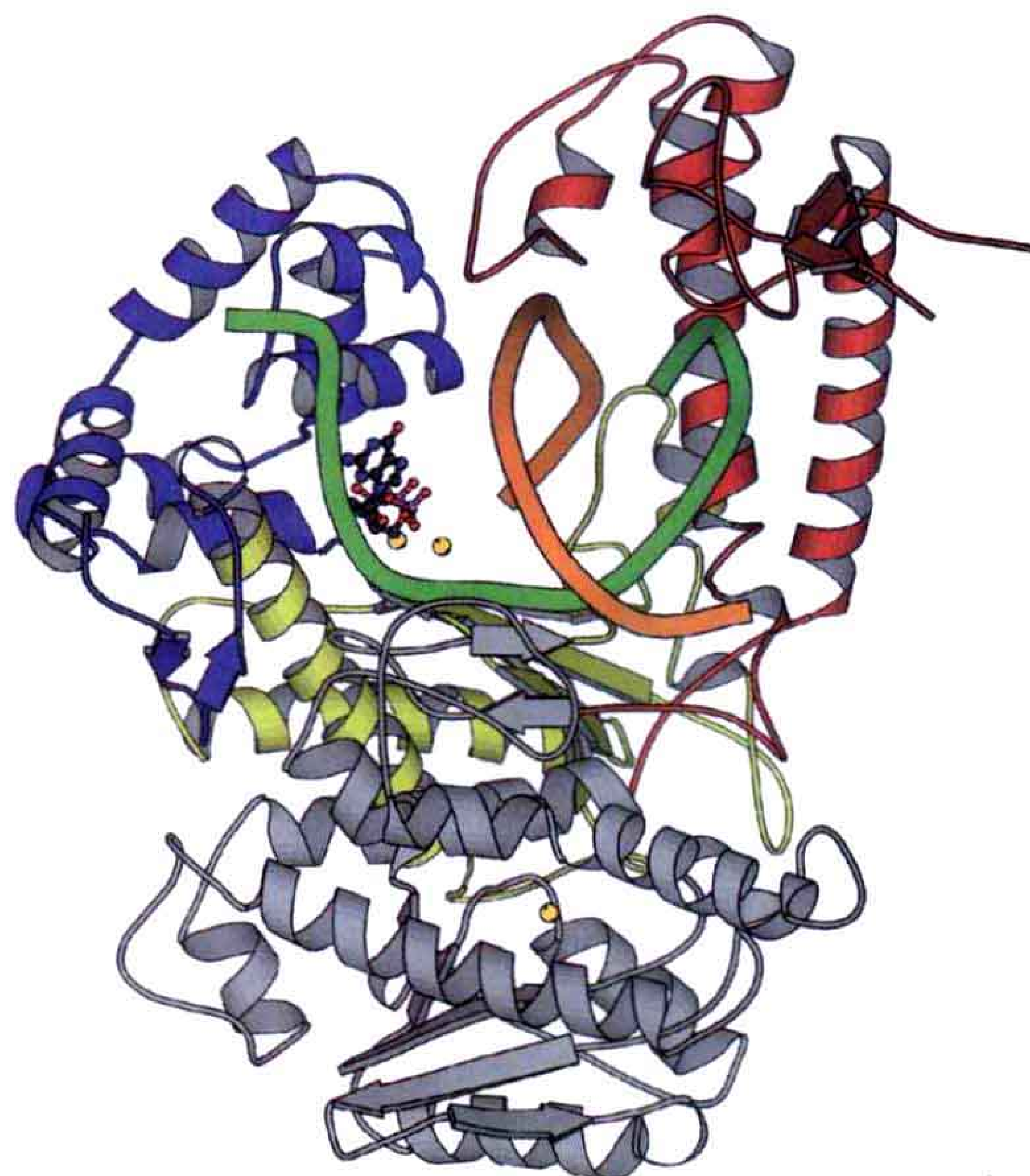


图 6.3 T7 DNA 聚合酶的结构。该结构可被视为一个右手结构，手掌显示为黄色，拇指显示为红色，四指显示为蓝色，参与催化反应的镁离子显示为黄色。Exo 结构域（灰色）参与校对。DNA 引物和模板分别显示为橙色和绿色。

除了 D 类 DNA 聚合酶，所有其他类的 DNA 聚合酶的结构也有被解析的，它们都含有 Klenow 片段的基本形状（图 6.4）。A、B、Y 和反转录类 DNA 聚合酶，以及病毒 RNA 复制酶都含有一个保守的手掌结构域，但是四指和拇指结构域并不保守，而且这些结构域在一级序列中也不是连续的。虽然大体结构相同，但是 C 和 X 类聚合酶手掌结构域的拓扑构象和在一级序列的位置却很不一样（图 6.5）。

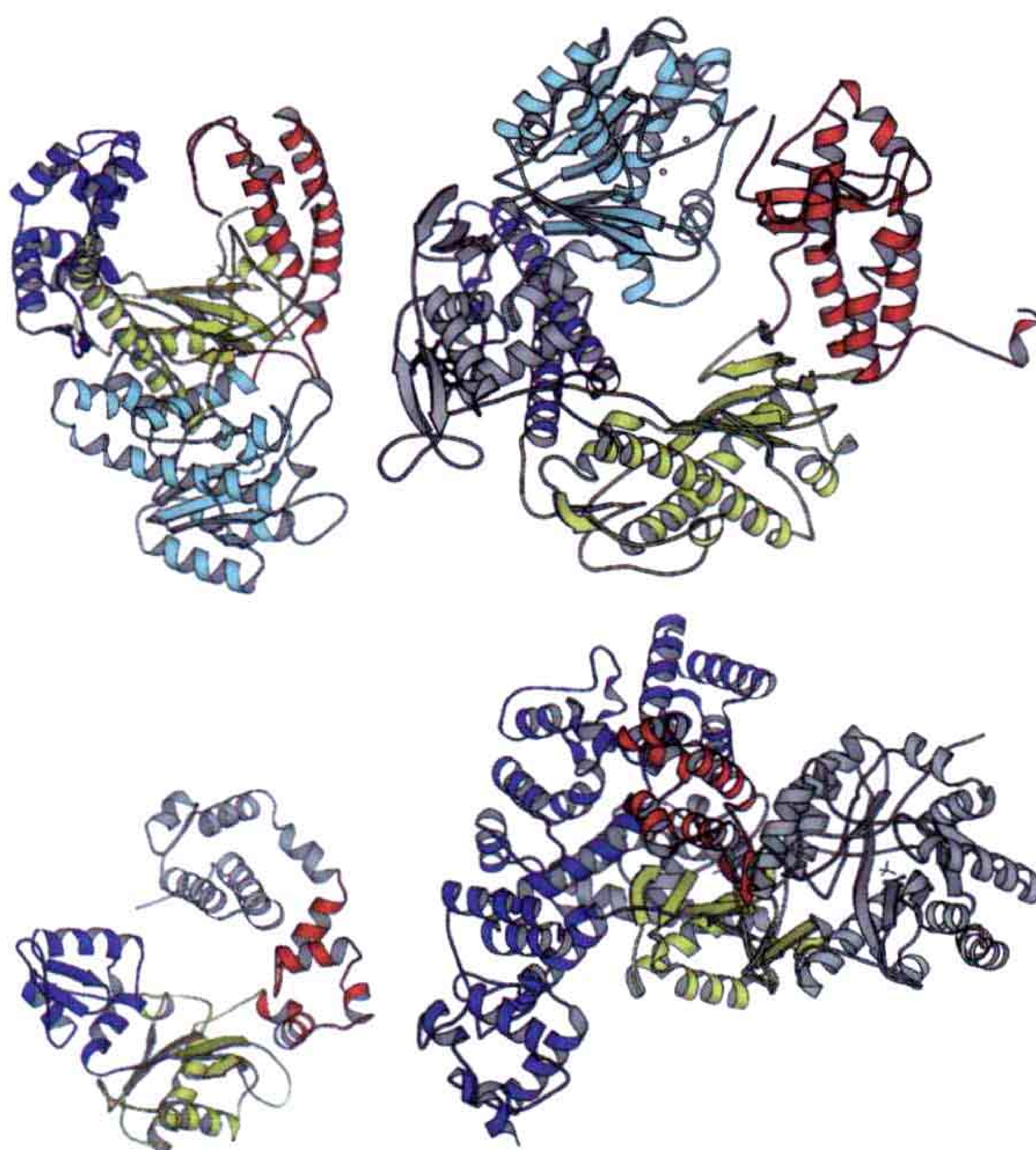


图 6.4 A、B、X 和 C 家族 DNA 聚合酶的结构。左上：Pol Ⅰ，Klenow 片段（A 类，PDB：1KFS）；右上：噬菌体 RB69 的 DNA 聚合酶（B 类，PDB：1IH7）；左下：Pol β（X 类，PDB：1PBX）；右下：Pol Ⅲ α 亚基（C 类，PDB：2HNH）。手掌结构域为黄色，拇指结构域为红色，四指结构域为蓝色。上图的两个聚合酶含有一个核酸外切酶结构域（青灰色）。这些结构域结构相似，但是其在相应分子中的位置有很大的差异。某些聚合酶还含有另外的结构域，用灰色显示。

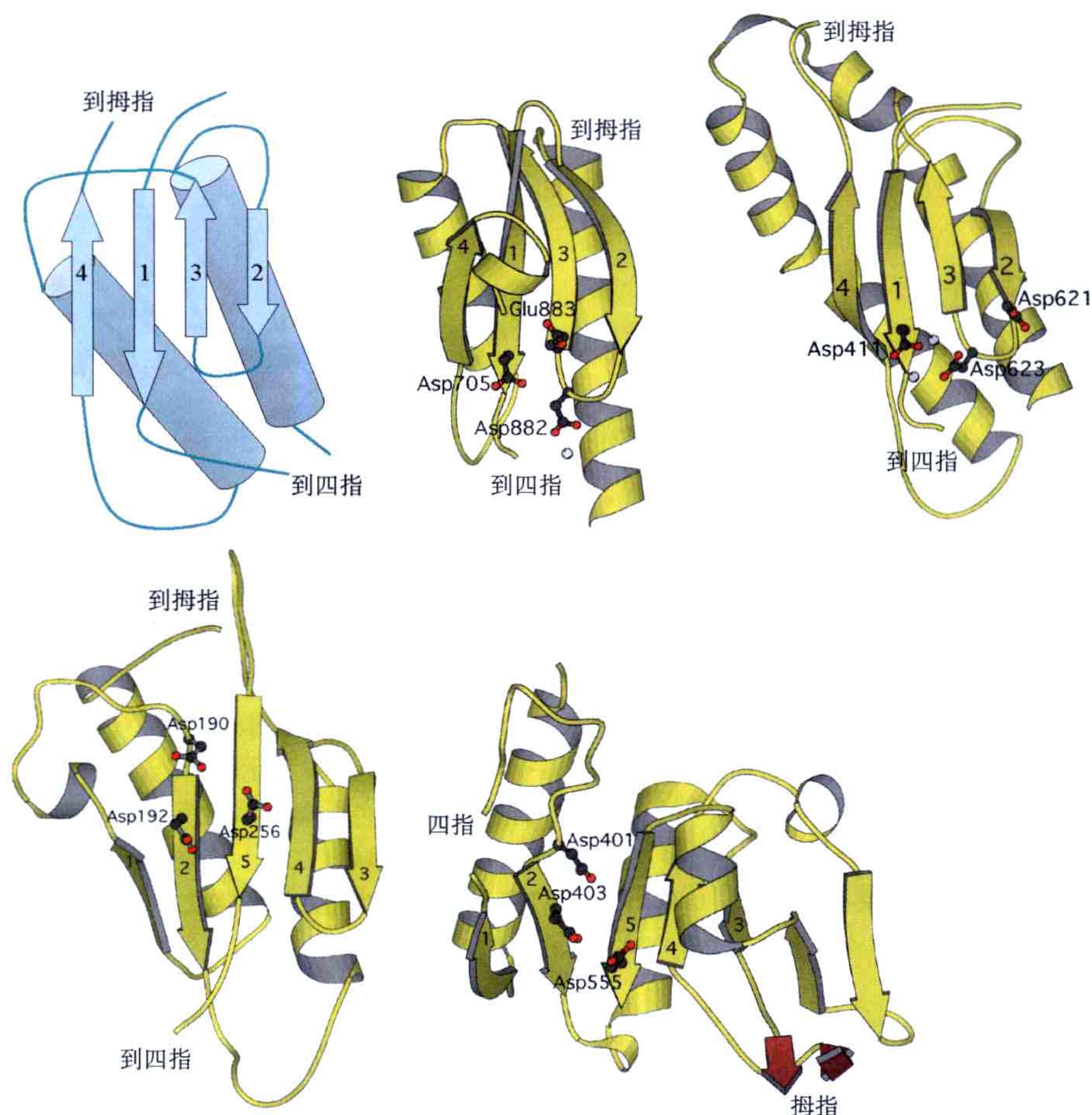


图 6.5 左上：DNA 聚合酶的手掌结构域具有一个存在于很多核酸结合蛋白中的保守的双股分裂的β-α-β 折叠结构。Pol Ⅰ（中上）和 RB69 聚合酶（右上）与 Pol β（左下）和 Pol Ⅲ（右下）的拓扑结构不同。参与结合镁离子的三个天冬氨酸或谷氨酸在图中标出。

DNA 聚合反应循环中的初始事件是 DNA 与聚合酶的结合，生成酶 - 引物 / 模板复合体（enzyme-primer/template complex）。几个这种复合体的晶体结构已被解析。细菌中的 Klenow 片段，形成复合体时主要的结构变化发生在拇指亚结构域，位于拇指尖的氨基酸残基向 DNA 分子靠近（图 6.6）。这种构象变化发生在一个螺旋 - 环 - 螺旋模体（helix-loop-helix motif）中，DNA 的结合使得该区域由无序的柔性结构转变为有序结构。

实际上，在整个 DNA 聚合反应过程中，会围绕 DNA 形成一个圆柱状结构。手掌与 DNA 的小沟及引物的 3′ 端相互作用。单链 DNA 模板在其骨架上形成一个锐角，从而使其与新合成的 DNA 双链位于拇指与四指间的缺口同侧，而且模板和 DNA 双链都不穿过缺口或圆柱体结构。一个酪氨酸残基（Klenow 片段的 766 位）堆积在新合成的 DNA 双链第一个碱基对的模板碱基上，将 DNA 置于可以插入新的核苷酸的合适位置上。

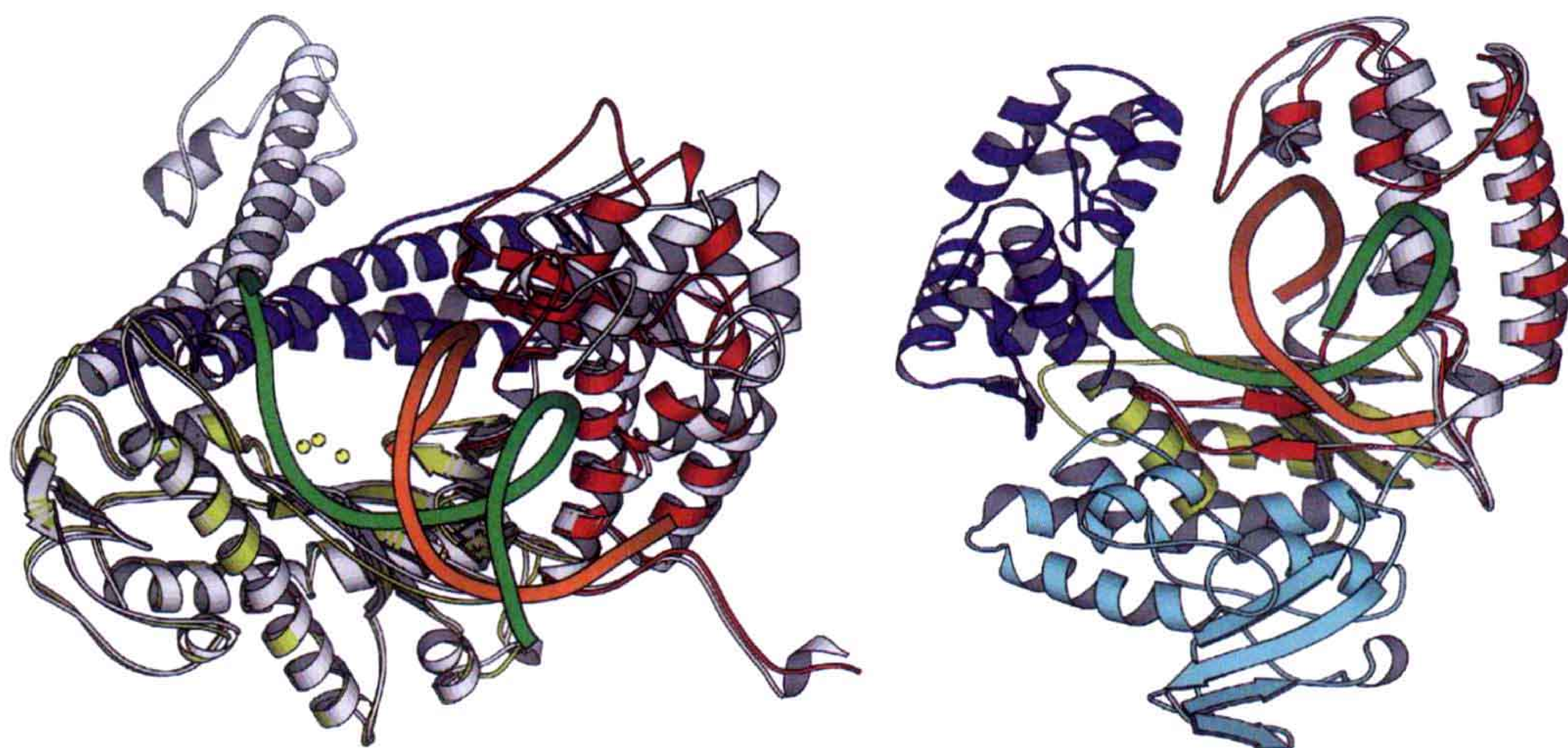


图 6.6 DNA 聚合酶结合模板发生的构象变化示例。左：噬菌体 RB69 的聚合酶，拇指结构域向核酸方向轻微移动，同时四指结构域发生一个巨大的构象变化使参与催化的氨基酸残基向新插入的核苷酸移动（金属离子以黄色显示）。未结合底物的空酶（apo）显示为灰色。N 端结构域和外切核酸酶结构域（PDB：1IG9 和 1IH7）在此没有显示以便观察。右：栖热水生菌（*Thermus aquaticus*）聚合酶 I 的 Klenow 片段，拇指结构域及其指尖向 DNA 双链方向弯曲。未结合底物的空酶显示为灰色，其他结构域结构与其相应的 DNA 复合体结构相似，在此没有显示（PDB：4KTQ 和 5KTQ）。

第二步是核苷酸的插入，由 dNTP 与酶 - 引物 / 模板复合体的结合起始。在这一步中，DNA 聚合酶可以区分而不结合错误的位置和错误的核苷酸。dNTP 与几个聚合酶复合体结构的解析，揭示了该步反应的限速步骤与手指结构域的关闭有关。不同的聚合酶结合正确的和错误的核苷酸的效率差别很大，而这又似乎决定了聚合酶的功能。参与复制的 DNA 聚合酶的选择性通常要比参与修复的聚合酶的选择性大很多，也重要很多。

来源于不同物种的结合了所有核苷酸以及被称为 O 螺旋（O-helix）的四指亚结构域的 Klenow 片段已经被结晶。核苷酸部分与 O 螺旋平行，并与带正电荷的精氨酸和赖氨酸相互作用。核苷酸的碱基指向 DNA 结合裂隙（DNA binding cleft）。核苷酸的选择性结合首先是通过磷酸的结合实现的，碱基的选择随后发生。结合在 O 螺旋上后，核苷酸距位于手掌结构域的活性位点仍有 10~15Å 的距离，所以手指结构域内需要发生很大的构象变化将核苷酸运送至活性中心。

B 家族的聚合酶可能具有一种不同的识别核苷酸的起始机制。人类的聚合酶 α 的核苷酸结合位点距离活性位点很近，位于模板的碱基处，可以通过碱基配对直接控制插入的核苷酸是否正确。

核苷酸插入过程的限速步骤是将酶 / 引物 / 模板 :dNTP 转化成活化的复合体。这一步对于磷酰基转移反应是至关重要的。在 A 家族的聚合酶中，该步反应在区分正确与错误的核苷酸中也发挥着重要作用。对于 T7 聚合酶和 Klenow 片段的熄灭流程（quench flow）的研究表明，插入正确和错误的核苷酸的速率比相差 2000~5000 倍，所以一旦有错误的核苷酸插入将大大减慢反应的进行。对于 Klenow 片段来说，一般

认为，在错误插入过程中聚合酶也发生构象变化，但是随后的化学反应决定了其较慢的反应速率。

换句话说，当结合正确的核苷酸时，聚合酶发生构象变化时的活性中心处于正确位置，从而促使随后的化学反应的发生（图 6.7）。而当结合错误的核苷酸时，聚合酶也发生构象变化，但是活性位点不能形成合适的几何构象，这大大减缓了化学反应的进行。

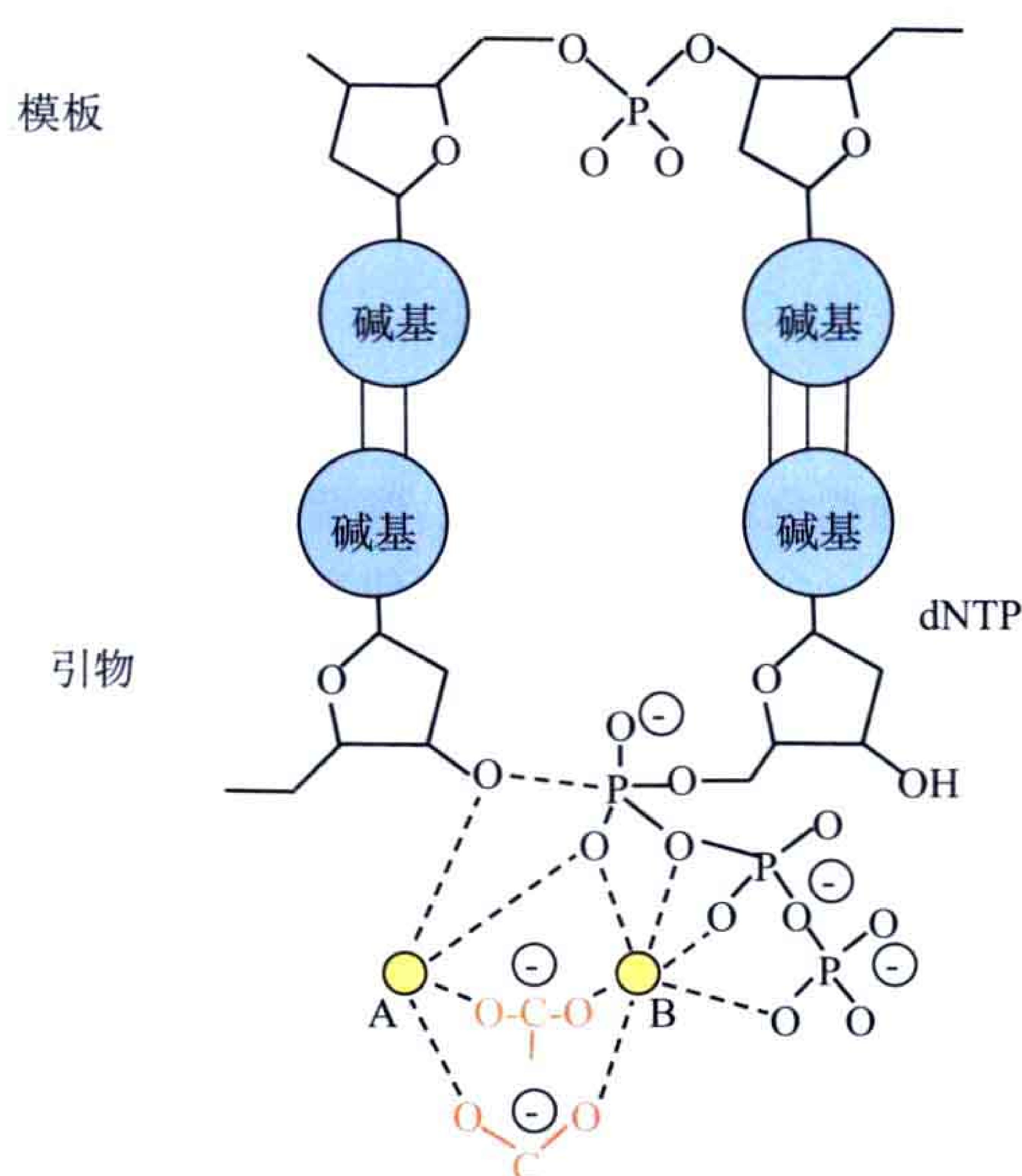


图 6.7　DNA 聚合酶位于手掌结构域的活性位点，在活性位点中含有两个镁离子 A 和 B（黄色），参与催化根据模板插入新核苷酸。这两个镁离子结合在两个天冬氨酸残基（红色）上。镁离子 A 活化引物的 3′-OH，用来攻击 dNTP 的 α- 磷酸。镁离子 B 结合在 dNTP 的 β 和 γ 磷酸上，稳定其负电荷 [该图来自 Brautigam CA, Steitz TA. (1998) *Curr Op Struct Biol* **8**: 54-63，并获得允许进行修改]。

磷酰基转移反应后，聚合酶发生第二个构象变化，释放产物焦磷酸（PP_i）。引物 / 模板的移动与手指亚结构域的打开同时发生。在手指亚结构域处于打开的状态下，DNA 分子可以沿着由保守的带正电荷的氨基酸残基构成的 DNA 结合位点的静电通道移动。这样，新合成的碱基对可以很快地在开打的复合体中移动，从而可以起始下一轮核苷酸的插入。

6.1.2.2　DNA 聚合酶的持续合成能力

在复制过程中，DNA 聚合酶需要成千上万次重复催化同样的化学反应，因此如果聚合酶可以保持与模板的结合，催化反应的效率就会比较高。这在很多情况下由聚合酶本身完成，而在另一些情况下需要一个单独的被称为持续合成因子（processivity factor）的蛋白质参与。一些持续合成因子呈环状或滑动的嵌状用以包围核酸分子（图 6.8）。

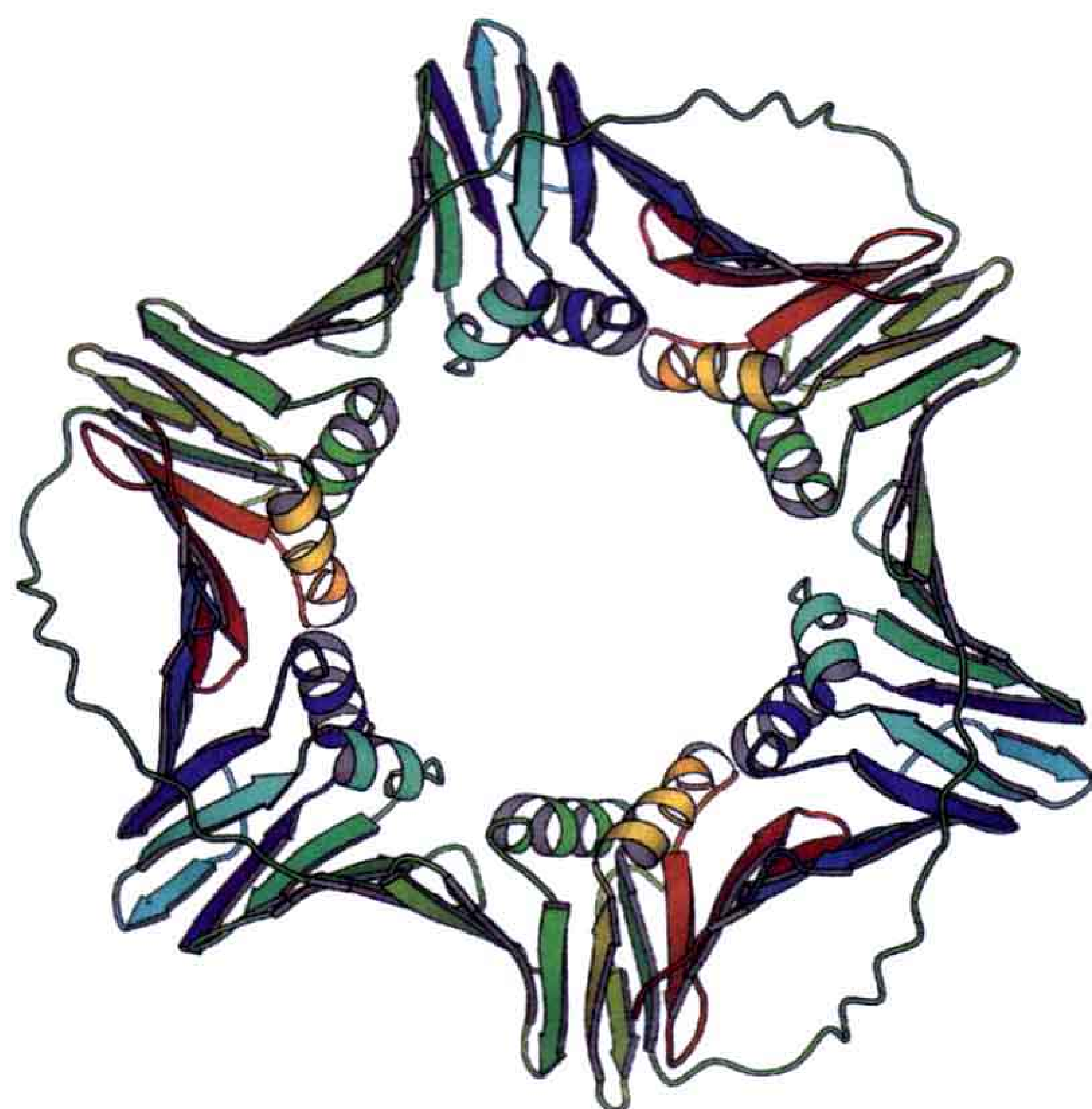


图 6.8 “滑动的钳子”，人类 PCNA (proliferating cell nuclear antigen)，用于包围新复制的 DNA (PDB: 1U7B)。

6.1.3 端粒酶

由于线性染色体的末端不仅需要被复制同时也参与保护 DNA 不受降解，所以线性染色体的复制方式与环形 DNA 复制方式有所不同。DNA 聚合酶只能沿模板的 3′ → 5′ 方向合成新的 DNA 链（图 6.9），这种方式对于合成引导链不存在问题，但是合成滞后链时所形成的冈崎片段是不连续的，在染色体的末端没有足够长的模板以供合成冈崎片段。所以，这种方式合成的新链的 5′ 端是不完整的，这使染色体会变得越来越短，经过几代的 DNA 复制之后可能会造成基因缺失。细菌的线性基因组的 3′ 端会有一系列的反向重复片段，这些片段会被很多蛋白质识别从而完成丢失的 5′ 端的复制。

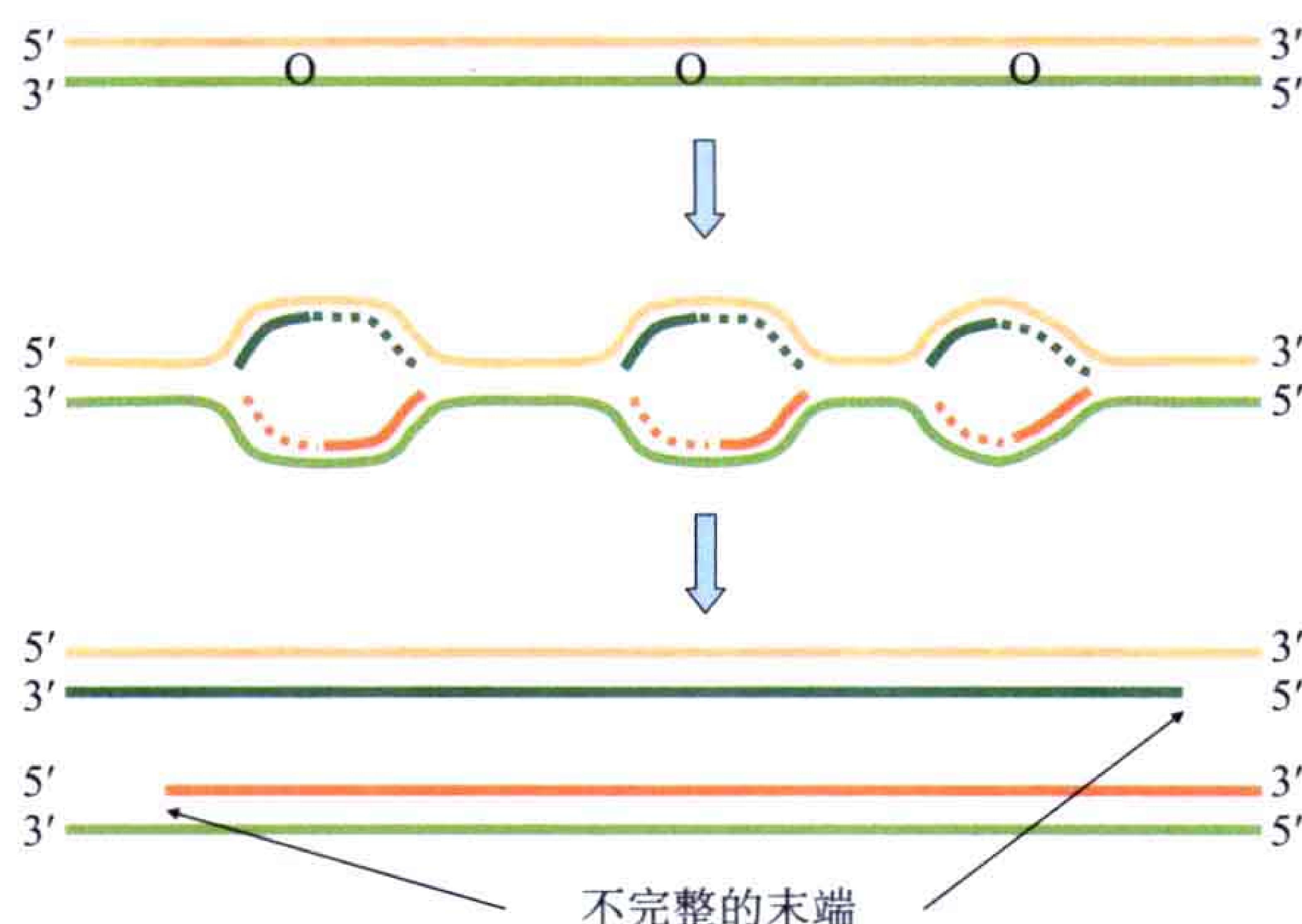


图 6.9 一段具有三个复制起点 O 的 DNA 的复制。在复制起点的前进方向，引导链被连续地复制，而滞后链则被合成为片段，这就造成合成的新链（较深颜色的链）5′ 端不是完整的。端粒酶协助补全合成的新链的末端。

真核生物染色体存在很多不编码基因也不直接参与基因表达调控的重要的 DNA 元件，其中包括 DNA 复制起点、参与染色体向子代细胞移动的中心体（centromer）及复制和保护染色体末端的端粒（telomere）。

端粒位于线性染色体的两端，包含很长一段结合了大量蛋白质的 DNA 单链。染色体末端为只有几个 bp 长但是重复成百上千次的重复序列。一个典型的重复片段长 6bp，富含 TG，如 TTAGGG，可以形成四链结构（见 3.4.4 节）。这些序列使染色体末端延伸了几千个 bp，蛋白质结合在这些序列上可以区分染色体末端和染色体断裂位点，从而保护染色体末端不参与重组和降解，这在很大程度上保护了染色体的结构不受重大损伤。另外，端粒作为特殊的复制起点使细胞可以复制染色体的末端。参与端粒复制的特殊的 DNA 聚合酶被称为端粒酶（telomerase）。

端粒酶是一个 1000 kDa 的核酸蛋白复合体，具有反转录酶活性（reverse transcriptase activity，TERT），将 RNA 编码的信息转化给 DNA。RNA（TER）是用于反转录 DNA 的至关重要的模板，从而延伸染色体末端的不能在正常复制循环中被复制的单链 DNA。TER 作为一个“记忆”元件（其序列与染色体末端单链 DNA 互补），可以一次次地“营救”染色体的末端。端粒末端核苷酸的 3′-OH 作为反转录的引物。TER 也含有几个具有二级结构的元件，可以和 TERT 相互作用以协助其插入正确的核苷酸、延伸端粒的重复序列、寡聚化及核定位（图 6.10）。除了 TERT 和 TER 之外，还有几个蛋白质也参与了端粒酶的组装、堆积、定位及其他功能。

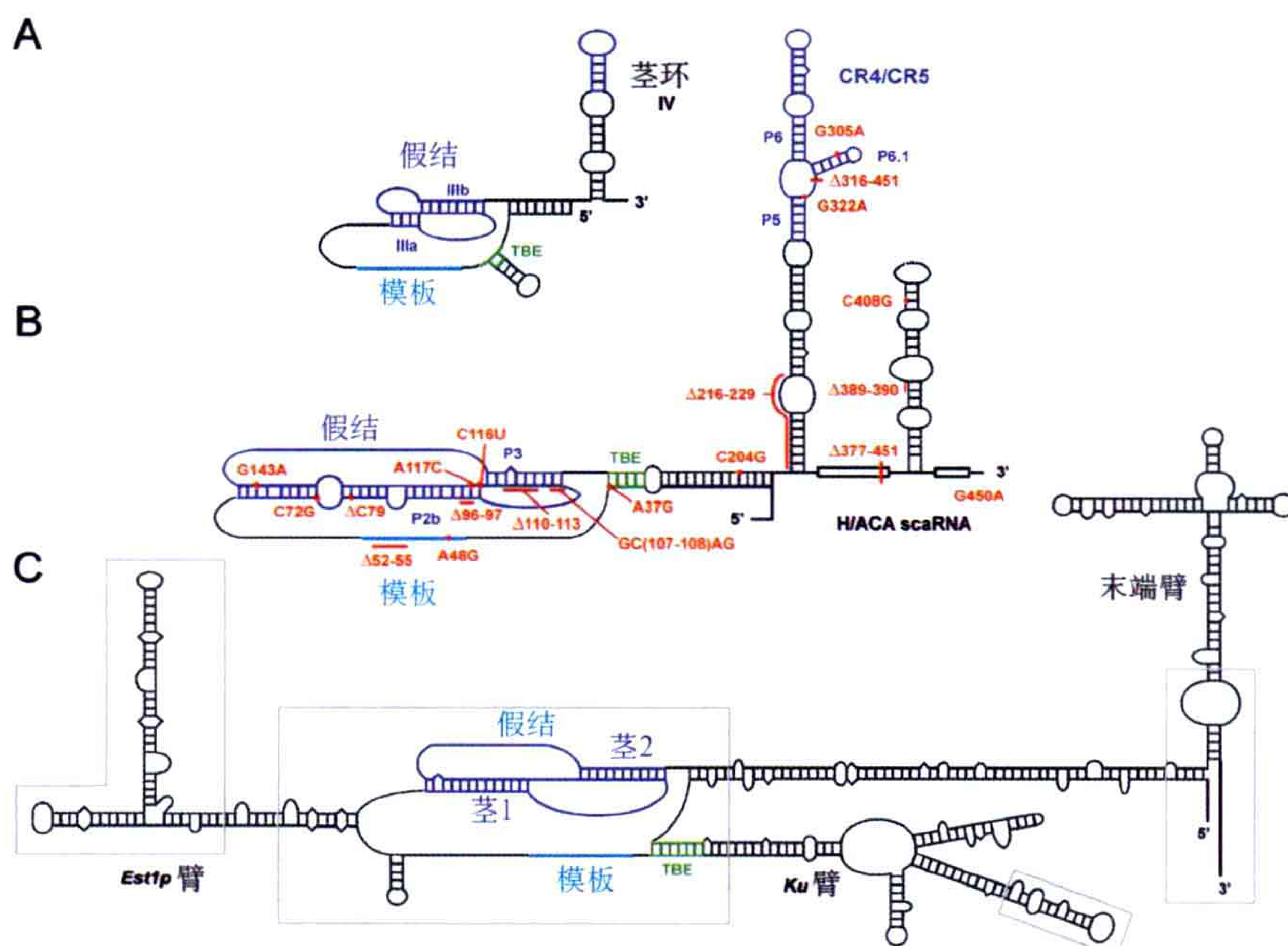


图 6.10 来源于纤毛虫（ciliate, *Tetrahymena thermophila*）（A）、人类（*Homo sapiens*）（B）和酿酒酵母（*Saccharomyces cerevisiae*）（C）的端粒酶 RNA（TER）的二级结构。模板显示为青色 [经允许引自 Theimer CA, Feigon J. (2006) Structure and function of telomerase RNA. *Curr Op Struct Biol* **16**: 307-318]。

在不同的物种中，TER 的一级结构和大小存在很大的不同，在纤毛虫中大约 150 个碱基，在脊椎动物中大约 450 个碱基，而在酵母中大于 1000 个碱基。然而，所有已知的 TER 都含有几个主要的元件：模板部分、5′ 模板边界元件（5′ template boundary element, TBE）和一个大环，这个大环实际包括模板部分、一个假结体和一个闭环螺旋区。在一些物种的 TER 中也存在其他的一些元件。

6.1.3.1　端粒酶全酶

来源于不同物种的 TERT 有几个共同的模体，同时也各自具有其物种特异的部分。简言之，其总体结构与其他核酸聚合酶的“右手”模型相似（图 6.11）。中心区的 RT 结构域（RT domain）含有 7 个通用的反转录元件。一般来说，在分别位于手掌和四指结构域的两个相邻的模体 A 和 B′ 中间有很大一段插入序列，这个参与序列夹在四指结构域中。与其他 DNA 聚合酶相比，TERT 含有两个参与 DNA 和 RNA 结合的 N 端结构域，以及一个短的 C 端延伸。

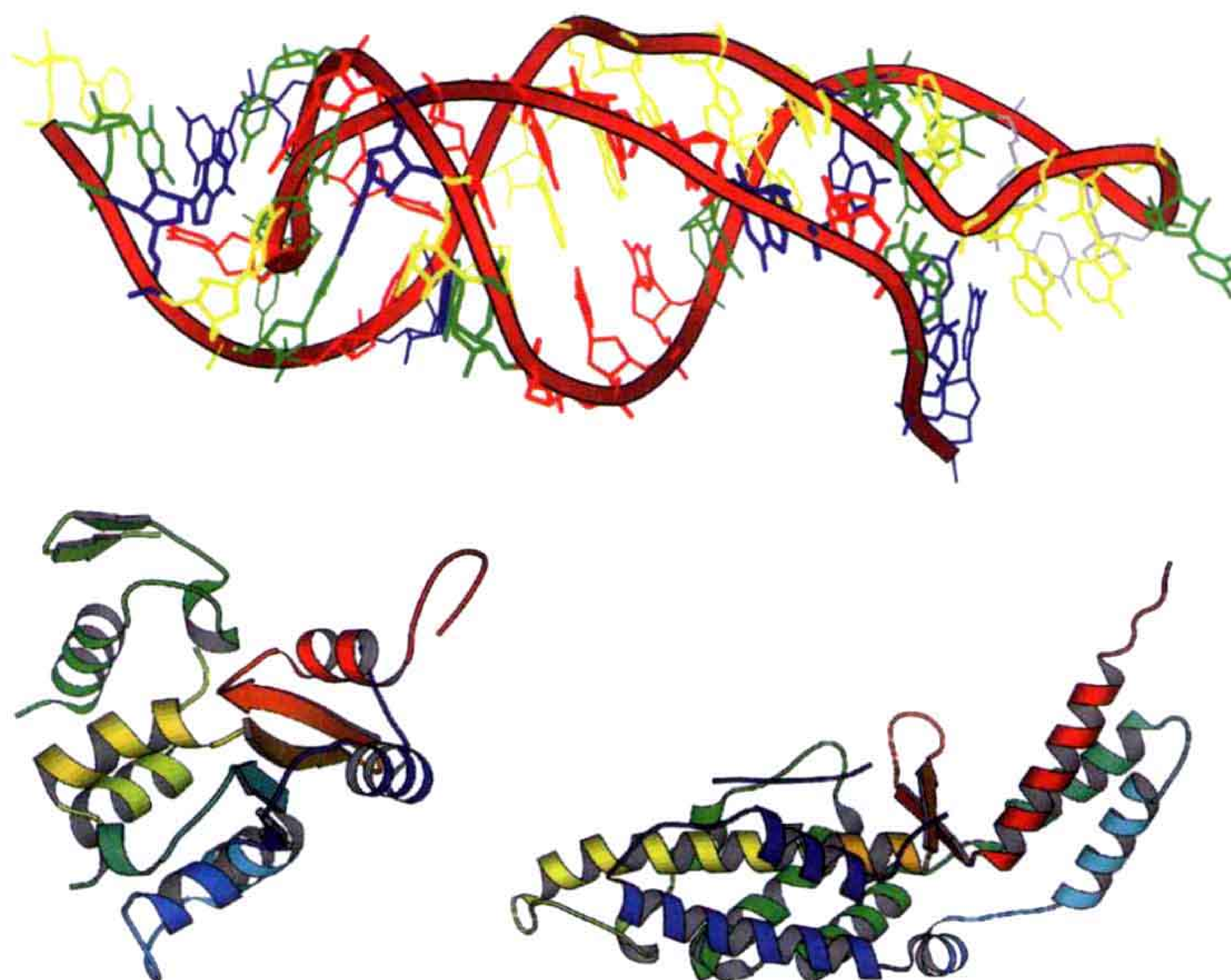


图 6.11　上：人类的 TER 假结体。其中的核苷酸来源于 stem p3 和 p2b（图 6.10）。鸟嘌呤、胞嘧啶、腺嘌呤和尿嘧啶分别显示为蓝色、绿色、黄色和红色。构成 RNA 分子的两个核苷酸显示为灰色。该分子中心部分有一个三螺旋结构（PDB：1YMO）。左下：来源于纤毛虫（*T. thermophila*）TERT 的 N 端 TEN 结构域。这个结构域含有一个参与 DNA 结合的独特的折叠结构（PDB：2B2A）。右下：来源于四膜虫（*Tetrahymena*）TERT 的 RNA 结合结构域，这个结构域与其他蛋白质没有很好的相似性（PDB：2R4G）。

仅 TER 和 TERT 即可重组端粒酶的活性。但是，在细胞中，端粒酶是一个含有很多其他蛋白质的巨大的复合体。原核生物纤毛虫（*Tetrahymena thermophila*）的基因组中含有几千个端粒，为研究端粒酶全酶复合体的结构提供了最初的线索。至少有 4 个另外的蛋白质 p20、p45、p65 和 p75 是全酶不可缺少的组分，它们在维持端粒的长度

中发挥作用，TER 则参与复合体的堆积和组装。

在细胞中，p65 似乎是组装的中心，它首先结合 TER，然后结合 TERT，接着是其他蛋白质。然而，由于 TERT 可溶性不好，体外重组实验一般是用 TERT 结构域而不是其整个蛋白质。

6.1.3.2 酶学性质

端粒酶与其他反转录酶工作机理类似，但是它具有一个独有的特征，即它利用的模板嵌在一个大的 RNA 分子中。首先，端粒酶全酶识别染色体的末端，染色体 DNA 的 3′ 端与 RNA 模板杂交配对；然后，模板直接在 DNA 的 3′ 端加入新的核苷酸直至达到模板的 5′ 端。随后发生一个独特的步骤，端粒酶转移至新合成的 DNA 的 3′ 端，然后发生另一轮的反转录。多步的转移和延伸反应形成了 DNA 末端的重复序列。

与其他 DNA 聚合酶类似，端粒酶催化的反应也依赖两个金属镁离子，因此位于活性位点的结合金属的天冬氨酸和谷氨酸是保守的。但是，在没有高分辨率全酶结构的情况下，很难推断反应机制。

6.1.3.3 其他解决方法

端粒酶存在于绝大多数真核生物中，是细胞的正常组分，它与由转座子编码的反转录酶起源相同。但是，起源相同的这些酶各自发展出不同的修复染色体末端的方法。例如，一些昆虫虽丢失了端粒酶基因，却发展出一种类似于反转座的机制用以保持染色体末端的完整性。染色体末端变短在老化和癌症发生中起到关键作用。因此，对于端粒酶结构与功能关系的更进一步的细节研究对医学应用意义重大。

6.2 重组和修复

在所有的细胞中，DNA 的结构总是在变化的。第一，DNA 长期暴露在物理、化学及生物试剂的“进攻”下，这些都会使 DNA 的双螺旋结构发生变化，这对细胞的基因构成造成威胁，需要修复。第二，DNA 总会发生重组，这给遗传物质提供了多样性。所有这些 DNA 的变化都是一系列过程的一部分，需要很多酶的参与，以保持 DNA 恒定和变化之间的平衡。同时，细胞必须一直检查 DNA 分子，并发现 DNA 损伤。在此，我们仅讨论一部分相关的酶。

在 DNA 修复过程中，被识别的损伤的 DNA 链会被切除并被新合成的链取代。下面以 oxoG 糖基化酶（oxoG glycosylase）为例讲述 DNA 修复的机制。即使是双螺旋损伤的 DNA，其损伤部分也可以被同源 DNA 的副本所取代。DNA 双链的修复与 DNA 重组有很大的重叠，这可以增加基因多样性。

6.2.1 DNA- 修复糖基化酶（DNA-repair glycosylase）

当细胞暴露在辐射和诱变试剂下时，这些环境因素会对 DNA 造成损伤。DNA 也会在水的作用下发生自发损伤，如胞嘧啶脱氨基和鸟苷氧化。

在有氧呼吸过程中，分子氧被还原成水会产生部分还原的中间体和其他副产物。这些自由基是潜在的亲电氧化试剂，它们会攻击细胞组分。DNA 是很易受自由基攻击的靶标，最常见的 DNA 损伤是鸟嘌呤被氧化成 7,8- 二氢 -8- 氧桥鸟嘌呤（oxoG）（图 6.12）。这种被修饰的碱基既可以与胞嘧啶配对，也可以与腺嘌呤配对，这使其成为一个基因突变的热点，将 G-C 对转变成 T-A 对。OxoG 的产生和作用是哺乳动物细胞自发转化及随后癌症发生的很多重要起因中的一种，因此细胞必须保护自己抵御这种变化。

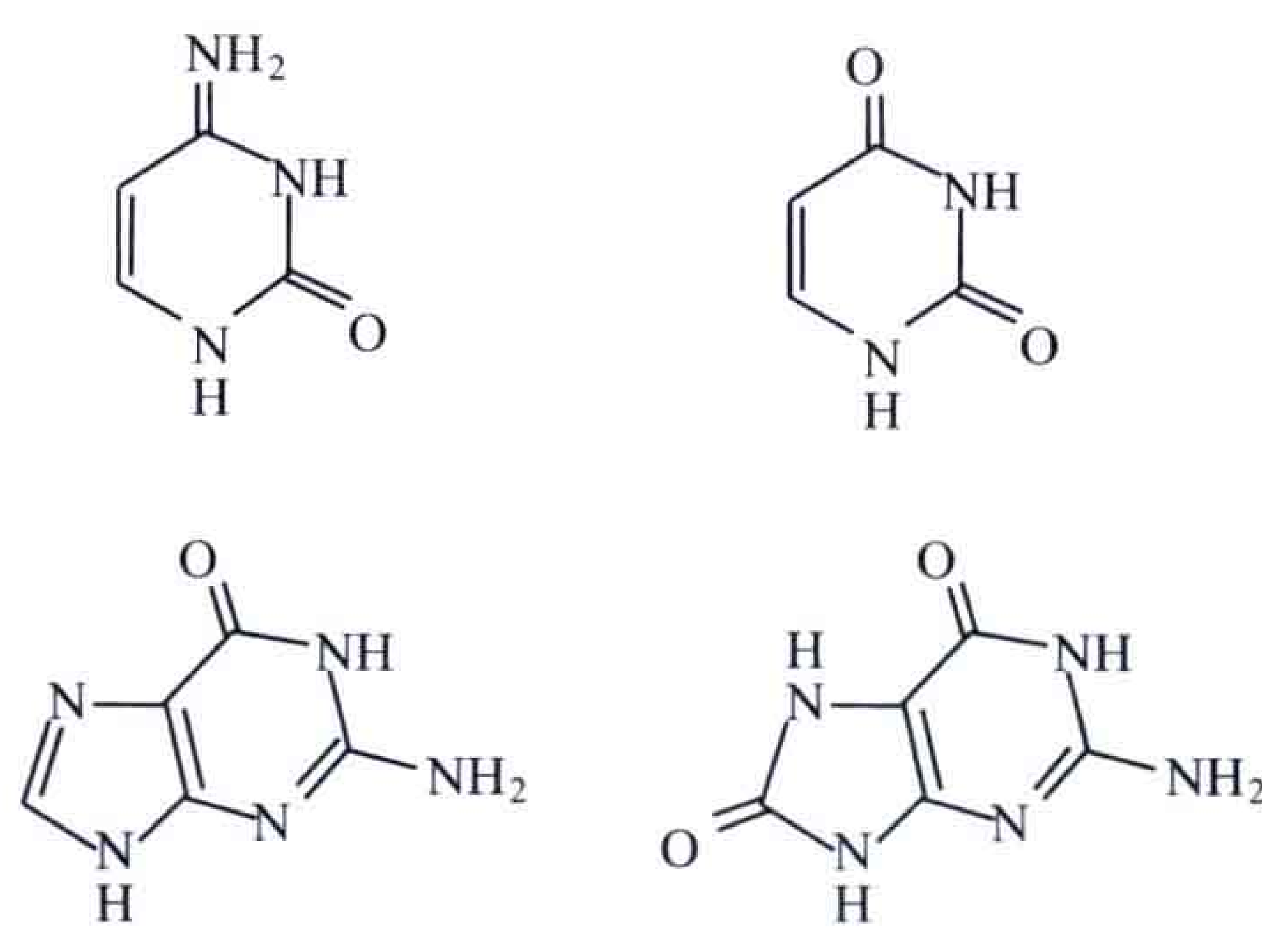


图 6.12 上：胞嘧啶脱氨基（左）生成尿嘧啶（右）；下：鸟嘌呤（左）氧化生成氧桥鸟嘌呤（右）

去除 oxoG 的一种方法是通过三磷酸酶系统的切除，通过识别 oxo-dGTP 并将其水解成 oxo-dGMP 而将其从核苷酸前体库中清除。然而，插入 DNA 中的 oxoG 是通过识别、清除和取代而被修复的。这种修复是由糖基化酶介导的，它识别这种被修饰的碱基并水解其糖苷键，剩下的脱氧核糖残基会被在 DNA 骨架上切除。然后，参与 DNA 修复的聚合酶和 DNA 连接酶会利用模板链将受损链修复完整。细胞中含有多种特异性不同的糖基化酶。人类尿嘧啶 -DNA 糖基化酶，负责将胞嘧啶脱氨产生的尿嘧啶清除。人类 oxoG-DNA 糖基化酶是第一个被深入研究的糖基化酶（图 6.13）。

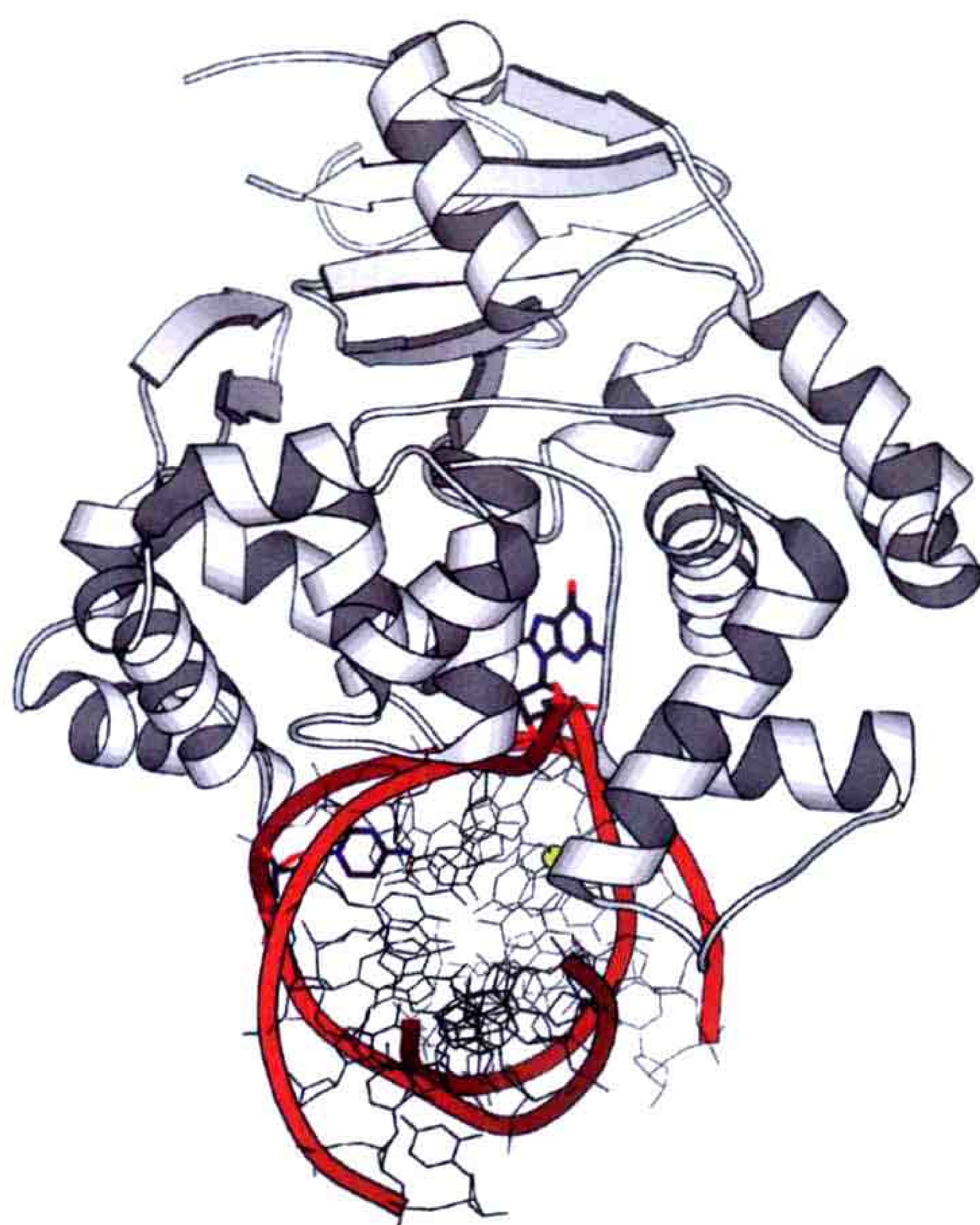


图 6.13 DNA 修复酶 oxoG-DNA 糖基化酶结合 DNA 的结构。oxoG 核苷酸已被从双螺旋中移出并结合在该酶的活性位点。与其配对的胞嘧啶用粗线显示（PDB：1EBM）。

哺乳动物 oxoG-DNA 糖基化酶由 *OGG*1 基因编码，是一个 DNA 糖基化酶 / β - 裂解酶，它识别细胞核或线粒体基因组中与胞嘧啶配对的 oxoG。破坏小鼠的 *OGG*1 基因会造

成 oxoG 在基因组中积累，增加自发突变率。人类表达两种不同的 oxoG-DNA 糖基化酶，一种定位在细胞核，另一种定位在线粒体。人类 *OGG*1 结合含 oxoG 的 DNA 复合体的结构为该反应机制所涉及的问题提供了很多答案。

如何找到埋藏在 DNA 双螺旋中的受损碱基和如何在其隐蔽的位置对其起作用是主要难题之一。该酶似乎在小沟中横向扩散扫描 DNA 直至发现受损碱基。DNA 双螺旋的柔性帮助 oxoG 碱基翻出双螺旋结构，直接结合在糖基化酶的活性位点。oxoG 的翻出仅仅适度地破坏了双螺旋结构，其能量消耗不大。在扫描 DNA 过程中糖基化酶是否需要将每个碱基翻出目前仍不清楚。

DNA 的修复还有其他几种机制。例如，受损的核苷酸（不仅是碱基受损）可以被识别并被从 DNA 骨架中去除，这种修复机制需要另外一套酶的参与，被识别的不是一个单独的被修饰的碱基而是双螺旋结构的扭曲。还有一种修复机制是基于同源重组的修复。

6.2.2 重组

同源 DNA 分子之间遗传信息交换是演化的主要过程之一。在细胞中，遗传信息交换一直都在发生，但是最极端的例子之一是减数分裂，它从根本上混合和重组染色体 DNA。真核生物中，分别来自父母的一对同源染色体在分离之前组合。遗传信息交换（genetic exchange, 也被称为 crossing over）发生在同源染色体配对过程中，这是同源重组最主要的事件。同源重组是由受调控的特定的酶催化的，对细胞至关重要。另外，重组可以找回 DNA 损伤丢失的序列，重起暂停的复制，并调控某些基因的表达。删除特定基因和引发定向突变等分子遗传学的中心方法都是基于同源重组发展起来的，如基因敲除小鼠。RecA 蛋白是参与同源重组的关键蛋白。

6.2.2.1 RecA

DNA 配对和链交换是同源重组过程的中心事件。在细菌中，RecA 是同源重组的中心酶，它负责寻找两个 DNA 分子中匹配的序列，并使两个分子中同源的部分发生配对。该酶属于链交换蛋白（strand-exchange protein）家族，几乎所有的生物中都存在与其相似的蛋白质。真核生物中，Rad51 发挥了相同的作用。

RecA 偏好结合单链 DNA，其单体与 DNA 相互作用发生聚合形成很大的纤维丝，可以与几条 DNA 链发生相互作用。RecA 聚合体的分子质量是可变的，可以有 100 个 RecA 单体参与形成有活性的复合体。这个复合体中包含一段松开的 DNA，为在不同的 DNA 分子中寻找同源序列提供可能，并进行 DNA 的交换。大肠杆菌 RecA 是第一个被解析晶体结构的 RecA 蛋白，由 Steitz 组于 1992 年解析。该蛋白质有一个主要的结构域，现被称为 RecA 折叠（RecA fold），它结合 ATP，和其他含有 P-loop 的蛋白质属同一类（图 6.14）。值得注意的是，在其他几种蛋白质中也存在 RecA 折叠结构域，

如解旋酶和 ATP 合酶的 F_1 部分（见第五章）。在很多晶型中，亚基以螺旋的形式存在，但是螺旋的程度有所不同。目前，尚未得到结合了 DNA 的复合体的晶体结构，所以关于 DNA 结合位点还需进一步研究和阐明。RecA 的 N 端和 C 端两个较小的结构域参与了 DNA 的结合和寡聚体结构的稳定。

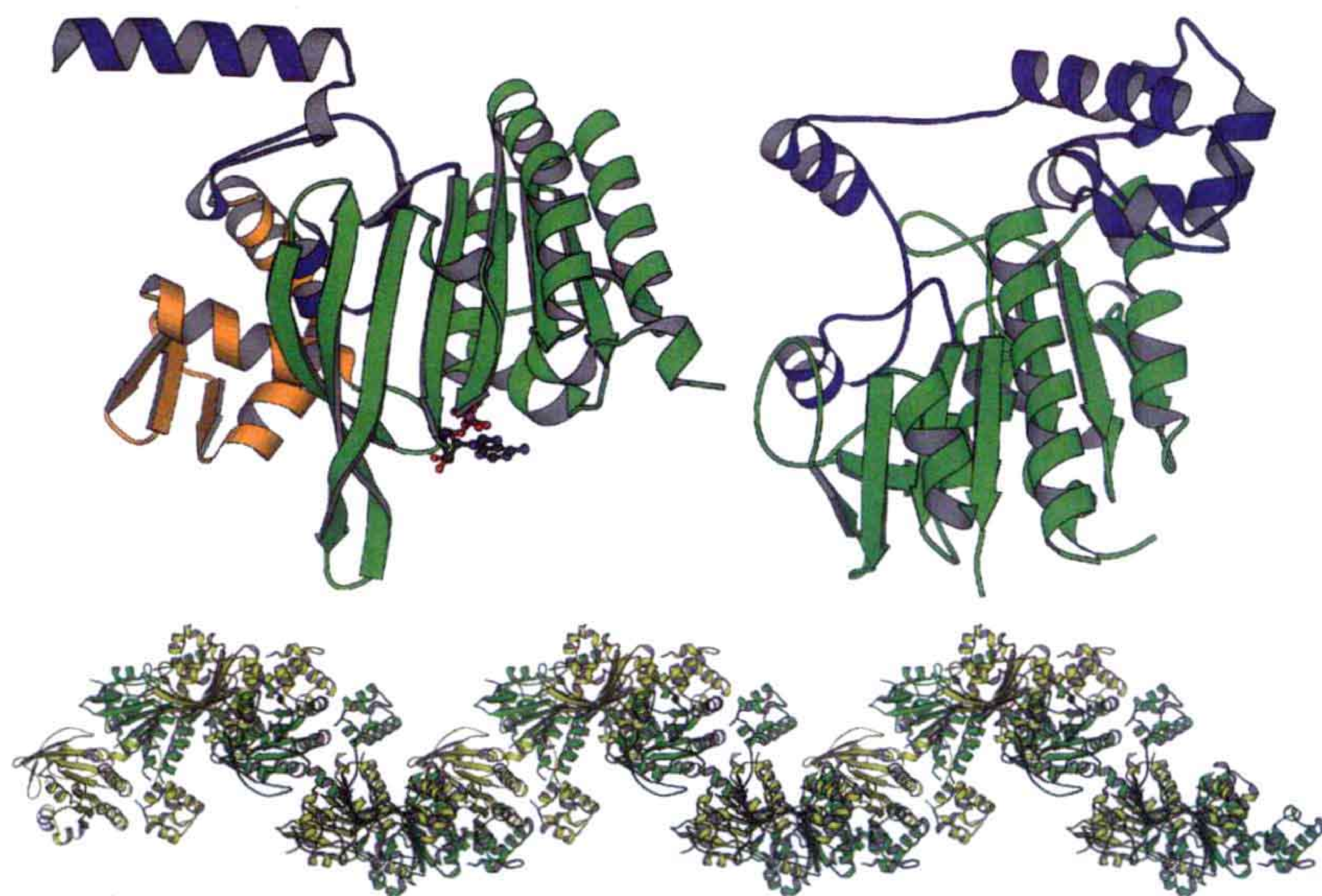


图 6.14　左上：大肠杆菌 RecA 的结构。其核心结构域显示为绿色，其 N 端和 C 端较小的结构域分别显示为蓝色和橙色，结合的 ADP 也在图中显示（PDB：1REA）。其核心结构域是很多蛋白质的 RecA 折叠的原型。右上：人类 Rad51 蛋白，它与 RecA 中心折叠的拓扑结构一样，但是其 N 端结构域不同（PDB：1SZP）。下：Rad51 的基于晶体结构的纤维丝状结构模型。它每个转角有 6 个亚基。该模型中蛋白质与蛋白质的相互作用与其功能研究相符。其与 DNA 的相互作用还未在结构中阐明，但是 DNA 应该结合在纤维的沟槽中。

延伸阅读（6.1 节）

原始文献

Beese LS, Derbyshire V, Steitz TA. (1993) Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**: 352-355.

Franklin MC, Wang J, Steitz TA. (2001) Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* **105**: 657-667.

Lamers MH, Georgescu RE, Lee SG, O, Donnell M, Kuriyan J. (2006) Crystal structure of the catalytic alpha subunit of *E. coli* replicative DNA polymerase Ⅲ . *Cell* **126**: 881-892.

综述文章

Autexier C, Lue NF. (2006) The structure and function of telomerase reverse transcriptase. *Ann Rev Biochem* **75**: 493-517.

Caruthers JM, McKay DB. (2002) Helicase structure and mechanism. *Curr Op Struct Biol* **12**: 123-133.

Hubscher U, Maga G, Spadari S. (2002) Eukaryotic DNA polymerases. *Ann Rev Biochem* **71**: 133-163.

Killoran MP, Keck JL. (2006) Sit down, relax and unwind: Structural insights into RecQ helicase mechanisms. *Nucl Acids Res* **34**: 4098-4105.

Rothwell P, Waksman G. (2005) Structure and mechanism of DNA polymerases. *Adv Prot Chem* **71**: 401-440.

Theimer CA, Feigon J. (2006) Structure and function of telomerase RNA. *Curr Op Struct Biol* **16**: 307-318.

延伸阅读（6.2 节）

原始文献

Bruner SD, Norman DPG, Verdine GL. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**: 859-865.

Parikh SS, Walcher G, Jones GD, Slupphaug G, Krokan HE, Blackburn GM, Tainer JA. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: Conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc Natl Acad Sci USA* **97**: 5083-5088.

Story RM, Weber IT, Steitz TA. (1992) The structure of the *E. coli* RecA protein monomer and polymer. *Nature* **355**: 318-324.

综述文章

Cox MM. (2007) Monitoring along with the bacterial RecA protein. *Nature Rev Cell Mol Biol* **8**: 127-137.

（王晓君 译　曹　骎　苏晓东 校）

第七章
转　　录

转录是细胞新陈代谢的重要过程，以 DNA 作为模板，利用基于 DNA 的 RNA 聚合酶（RNAP）从 4 种核糖核酸三磷酸合成 RNA 的过程。合成的方向是从 5′ 端到 3′ 端。RNA 是 DNA 模板链的一个直接拷贝，服从 Watson-Crick 碱基配对原则。因此，RNA 是 DNA 模板链的一个互补拷贝序列，它是一个 DNA 反链，或者编码的 DNA 链的精确拷贝，唯一的不同只是 DNA 中是胸腺嘧啶碱基，而 RNA 则换成了尿嘧啶。也存在基于 RNA 的 RNA 聚合酶，如在病毒中，只含有 RNA 基因组。RNA 分子合成后，主要是被用作细胞中的工具，如用在翻译中的 tRNA、核糖体 RNA、信使 RNA，或者控制基因表达的 RNAi 和小 RNA。转录的过程是高度受调控的，为了保证提供给生物体在不同的发育阶段和生理条件下需要的分子。

像所有的聚合反应一样，转录也可以被分为三个阶段：启动、延伸、终止（initiation、elongation 和 termination）。处于中心地位的酶是 RNA 聚合酶，用来把 DNA 复制成 RNA。为了拷贝 DNA，在启动阶段，最可能的开始是识别。RNA 聚合酶没法高效而特异地、独立地启动转录。这一过程因此需要一些蛋白质的帮助，这些蛋白质在启动阶段以激活子或者阻遏子身份出现。

7.1　DNA 中的控制原理

7.1.1　细菌的启动子

基因在编码区域之前和之后都含有特定的序列，其目的是使转录可以被控制。最主要的控制因子就是启动子，它是 DNA 上处在编码区域 5′ 端的一个片段。在细菌中，启动子是一段保守序列，出现在转录开始位点的上游，是 RNA 聚合酶的结合位点。在大肠杆菌中，启动子中有两个保守的位点，它们的中心分别位于 –10bp 处［普里布诺框（Pribnow box），保守序列为 TATAAT］和 –35bp 处（保守序列为 TTGACA），而

第三个位点则在更远的上游。激活因子和阻遏子都能结合到受控制基因的启动子区域，分别促进或者抑制 RNA 聚合酶结合到 DNA 上的转录基因。

7.1.2 真核生物中的启动子和调控原理

在真核生物中，一些定位在转录开始位点上游区域的序列对于控制转录非常重要。一个重要的序列就是 TATA 框，它定位在转录开始位点前大约 25 个核苷酸的位置。这一序列在很多但不是全部的基因中被找到。其他还有一些常见的序列，如 CCAAT 和 GC 框。然而，启动子序列同样可以出现在基因转录开始位点的下游。调控基因转录的系统中，还包含基因特异的转录因子，可以激活或者抑制基因转录的启动阶段，通过结合到控制元件上，其可以在基因的几千个碱基（kb）范围内被找到（图 7.1）。

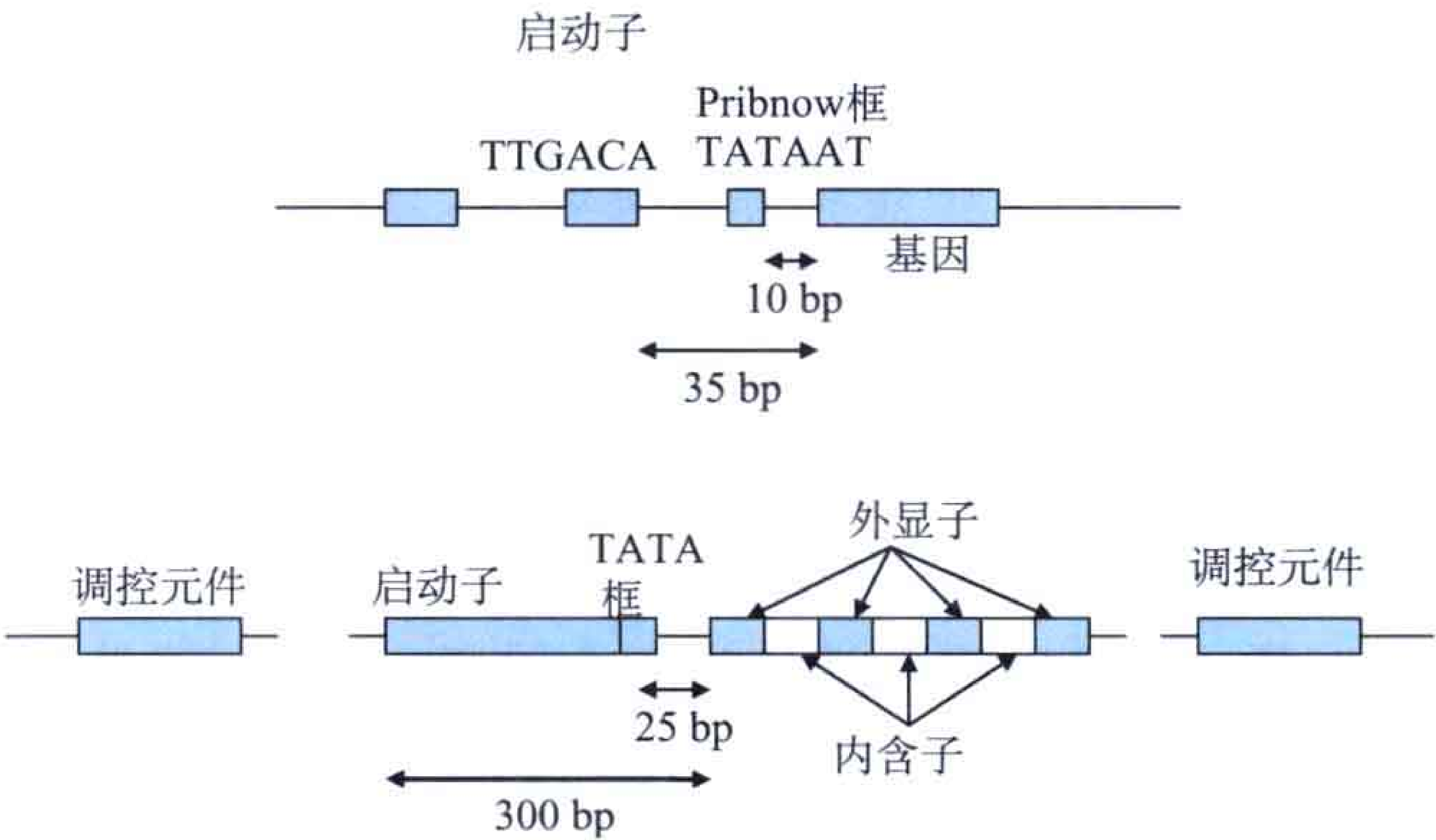


图 7.1 上：细菌基因的轮廓图。显示了这一基因的启动子，包括一些一般性的序列和特异性区域。下：真核生物基因的轮廓图。转录区域包含了内含子和外显子。内含子在转录后将被去掉。在转录部分的前边和后边不同的序列中包含了基因的控制信息。

7.1.3 通过 DNA 组件控制

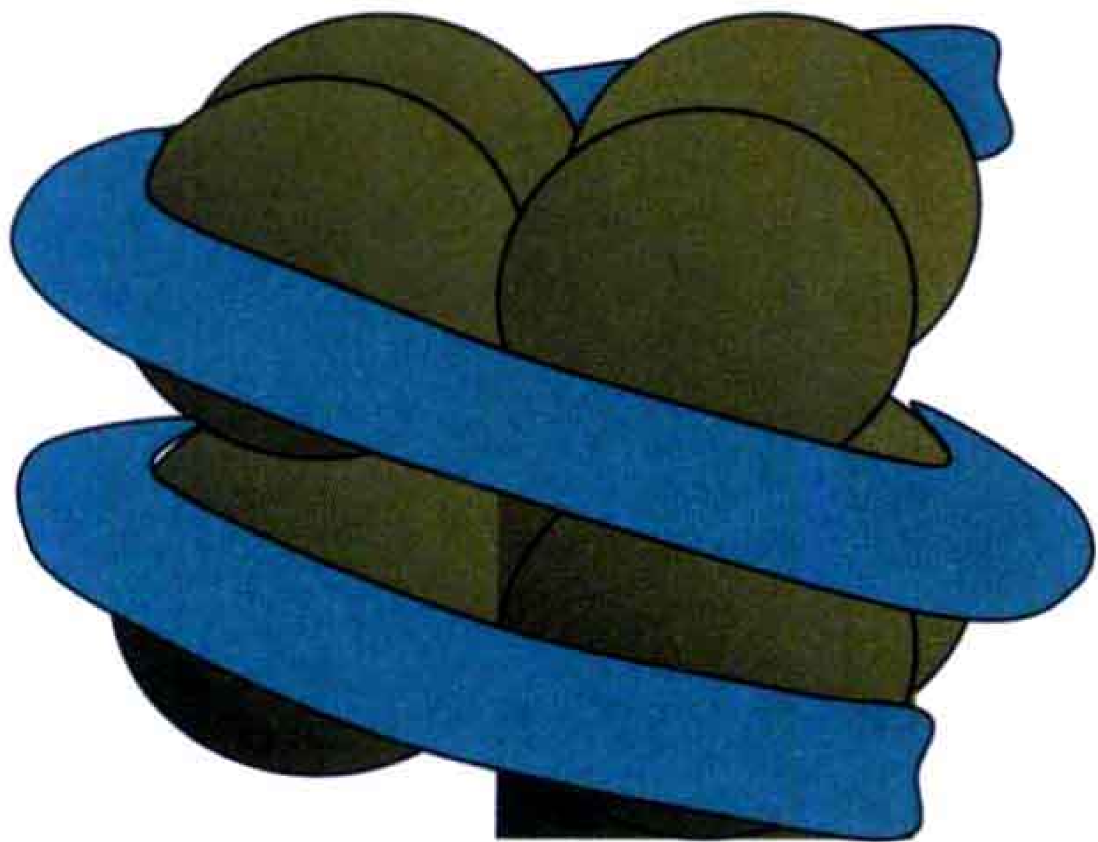


图 7.2 一个核小体的简单表示，中间是一个组蛋白八聚体（绿色），147 个碱基对长的 DNA（蓝色）以左手螺旋的方式缠绕在组蛋白上。

真核生物中生成 RNA 分子的过程要比细菌中复杂很多，它甚至通过基因组的其他组织来调控。DNA 被定位在细胞核内，同时紧缩成染色质，在染色质中，DNA 被缠绕在组蛋白八聚体上形成了核小体（图 7.2；见第三章）。这些核小体又会被组织成更大的结构。为了可以被转录，基因必须从核小体中释放出来。在真核生物中，这就是为什么需要更少的阻遏子来抑制基因转录的原因，因为核小体提供了一个很强的负调控因素。

从组蛋白的中心延伸出来的N端是柔性的，可以通过磷酸化、甲基化、乙酰化、泛素化和SUMO化被修饰（见附录E）。这些酶催化的修饰可以与激活子或者阻遏子形成基因特异性相互作用。这控制了核小体是否形成更高层次的结构，以便调控基因的表达。因为DNA需要暴露出来才能被转录，这些修饰，特别是组蛋白的乙酰化和脱乙酰化作用很强地调控了特异基因的转录。乙酰化一般都与转录有关，而相反的脱乙酰化与抑制有关。组蛋白尾部的赖氨酸在乙酰化以后，它的疏水性增强了，失去了一个正电荷。这使得它们与DNA的相互作用能力减弱了。组蛋白与DNA之间的相互作用的减弱，使得DNA更容易与转录因子及DNA聚合酶接触。

组蛋白的乙酰化作用不是可以使真核生物基因转录的充分条件。首先是启动子，但是后面的整个基因也需要处在可以被转录的环境下。一个蛋白质的区域通过基于ATP的机制被识别，涉及染色质的重构。这在很多机制中都有可能。当染色质的高级结构被解开，因为核小体的动力学性质，基因的控制区域可以有效地被暴露，或者成为可以接近的区域（图7.3）。在这一过程中，一些组蛋白也可能从颗粒中解离下来。另一个可以选择的方案是，轴心（组蛋白八聚体）可以在缠绕在它上面的DNA上滚动，使得先前被隐藏的启动子位点被暴露。对于染色质重建这个领域，我们期望结构生物学可以帮助我们更好的理解。

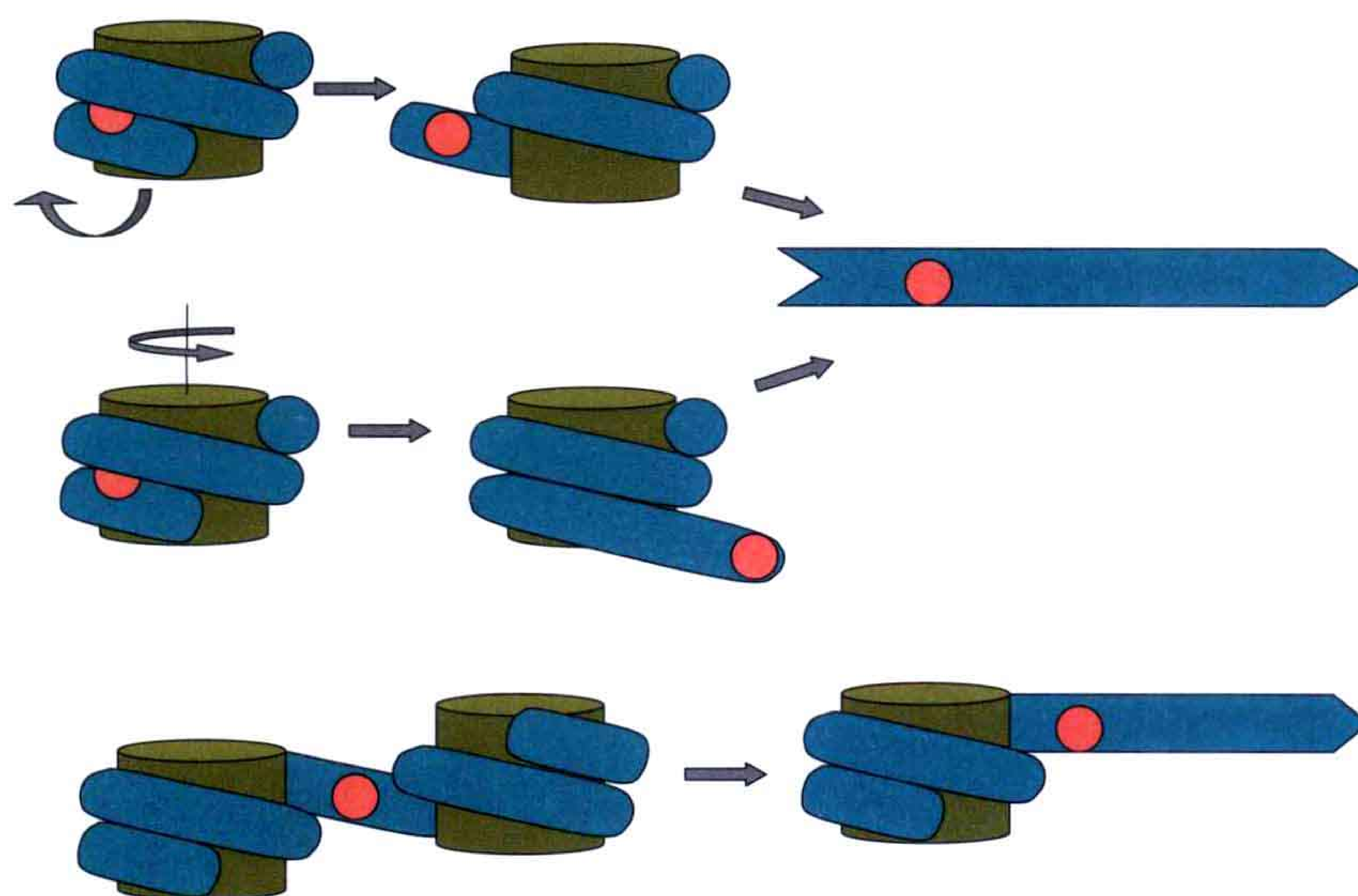


图7.3　基因的活性被核小体抑制。组蛋白八聚体由苔绿色表示，DNA由蓝色表示。上：核小体动态的短暂的松开使得其结合位点暴露给转录因子，之后这也导致了基因的暴露；中：因为核小体的移动，结合位点成为暴露状态；下：结合位点在核小体之间变得可以接近了。

一个关于组蛋白修饰很有意思的方向是它们可以通过细胞分裂把它们的信息传递给两个子链。组蛋白的修饰就是这样被遗传的。这成了一个单独的编码，超出了DNA中包含的信息，这就是所谓的组蛋白编码（histone code）。这是表观遗传学的一部分。一个基因也可以由于DNA中胞嘧啶碱基的甲基化而失去活性，这也是可以被遗传的特性。

7.2 细菌的转录

在细菌中，只有一种 RNA 聚合酶，而在真核生物中有三种。细菌酶的 5 个亚基是由 4 种类型的 4 个亚基（$\alpha_2\beta\beta'\omega$）组成了一个 RNAP 的核，第五个亚基，即 σ 或者 sigma，参与了启动子的识别并且具有转录的启动功能。RNAP 包括 σ 被称为全酶。这一复合物的分子质量为 400~450kDa。细菌激活子和阻遏子（见 7.6 节）与 RNA 聚合酶直接作用。

双链 DNA 需要被解旋后才能与 RNA 聚合酶相互作用（图 7.4）。解旋的区域被称为“转录泡（transcription bubble）”，它含有大概 15 个碱基对。转录的 RNA 与 DNA 的模板链形成了一个杂合的螺旋，它的长度为 8~9 个碱基对。细菌、真核生物也是这样，它们的 RNA 聚合酶都有能力向前或者倒退移动。核苷酸三磷酸在结合状态时可以向前移动，而在降解 DNA 时就会导致回溯移动。

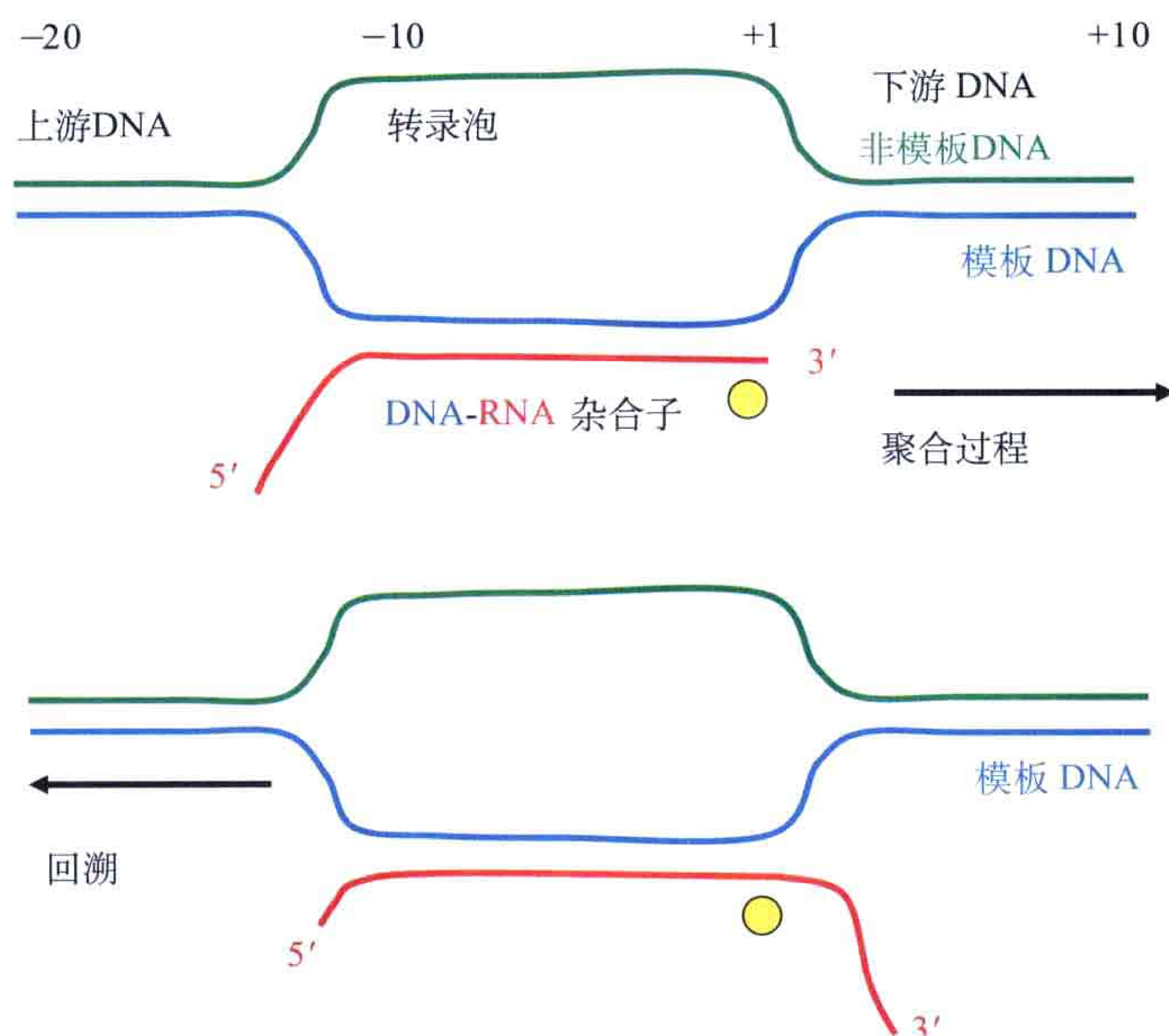


图 7.4 DNA 到 RNA 的转录需要一段双链 DNA 解旋的或者不缠绕的区域。这一区域是 RNA 聚合酶结合的位点，被称为转录泡（transcription bubble）。在转录中，转录泡中形成了 8~9 个碱基对的 DNA-RNA 杂合子。下游 DNA 还没有被转录时，上游 DNA 在它们被转录后就已经恢复了双链结构。RNA 聚合酶在 RNA 聚合过程中可以向前移动，但是在出现错误时，为了把错误的碱基给切除，回溯也是可能发生的。黄色环标记的是活性位点，这里是核苷酸加到 RNA 链上的地方。编号是关于聚合酶活性位点的而不是开始转录的基因。

在转录起始阶段，全酶结合到两个 DNA 启动子区域的六聚序列上，相对于通过 σ 启动子的帮助的转录起始位置（+1），这两个序列的中心分别在 − 35 和 − 10 的位置。这里有不同的 σ 亚基，可以找到不同的启动子位置。它们是细菌转录的主要调控因子。根据结合，全酶会在双链 DNA 的 − 12 和 + 2 位置解开螺旋，使得其形成一个开放的

启动子复合物，这样转录就能被启动了。当 9 个碱基对的 DNA-RNA 杂合子形成后，转录的起始阶段就结束了，而延伸阶段就随之开始。延伸复合物具有很高的稳定性，可以合成上千个核苷酸长度的 RNA。

7.2.1 细菌 RNA 聚合酶的结构

RNAP 的结构就像一个螃蟹钳子一样，它们各自的大小分别是大约 150 Å 长和 110 Å 长（图 7.5）。RNA 聚合酶的结构与 DNA 聚合酶具有显著的差别。双链 DNA 当进入 RNAP 的钳状结构后便开始打开，形成不缠绕的或者解旋的状态，然后再转录成 RNA 链。钳子的一半是由 β 亚基（124kDa）组成的，而另一半则是由 β′ 亚基（171kDa）组成。这些大亚基每个都是由很多结构域构成的。这两个亚基之间存在一个很宽的界面。β 亚基通过一个柔韧的侧翼来环抱 β′ 亚基。β′ 亚基的一个结构域也与 β 亚基形成了一个很相似的作用。β′ 亚基是环状的，它们的 N 端和 C 端互相靠得很近。α 亚基含有两个结构域，它们的 N 端结构域的二聚体出现在钳子的底部，进一步增强了它们的相互作用。β 亚基几乎唯一地与其中一个 α 亚基相互作用，而 β′ 亚基则与另一个 α 亚基相互作用。β 亚基和 β′ 亚基共同形成了活性位点。ω 亚基环绕 β′ 亚基的 C 端尾部，没有与活性位点接触。

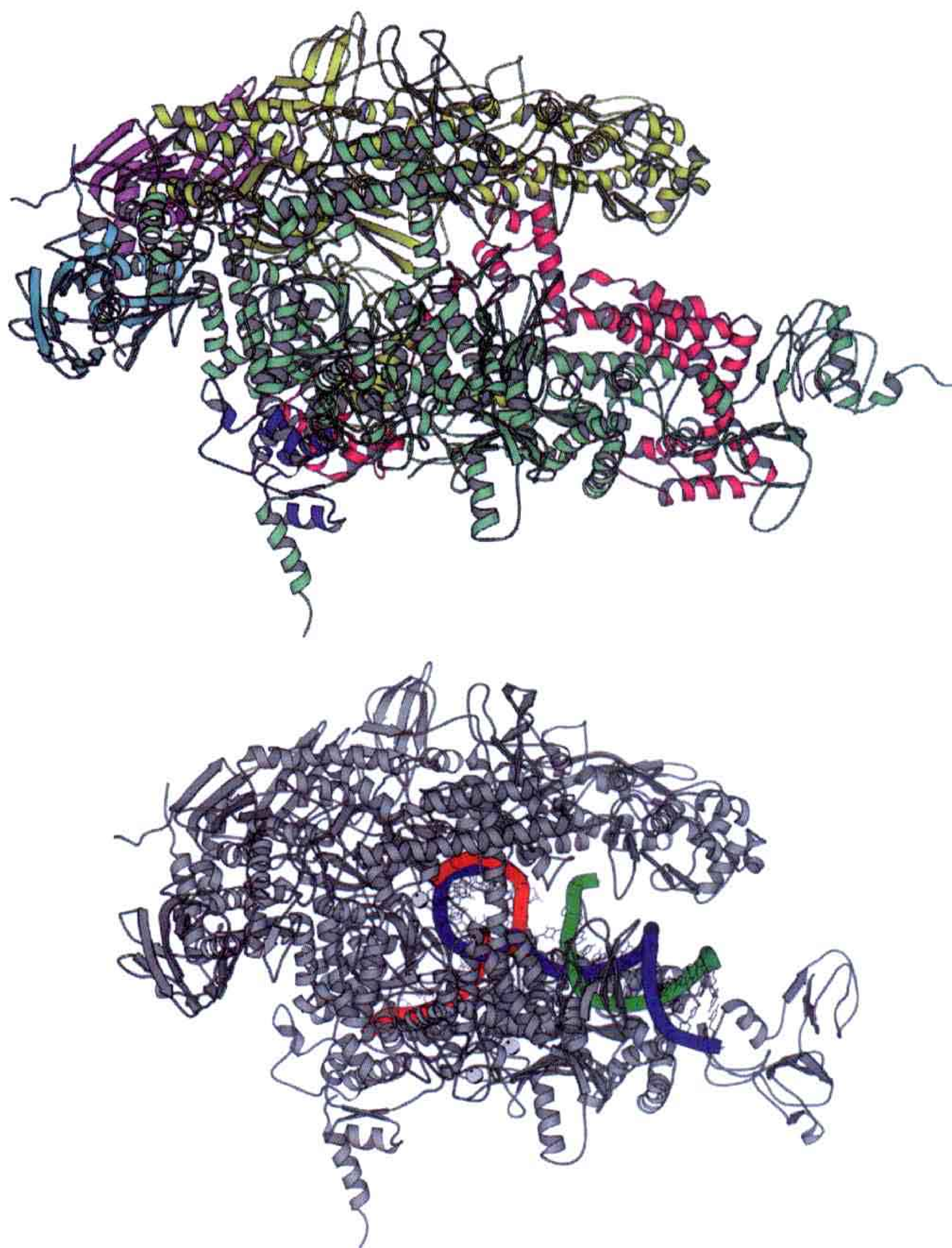


图 7.5 上：来自 *T. thermophilus* 的全 RNA 聚合酶的结构（PDB：1IW7）。结构主要部分的 β（黄色）和 β′（绿色）亚基看上去像一个螃蟹钳子。两个 α 亚基（Ⅰ 紫色和 Ⅱ 亮蓝色）、ω（深蓝色）和 σ（红色）都在外侧位置。下：细菌 RNA 聚合酶的结构，活性位点含有下游的 DNA 及 DNA-RNA 杂交体（PDB：2O5I）。非模板链 DNA 为绿色，而模板 DNA 则是蓝色的，RNA 链为红色。

活性位点被确认位于裂口的底部，通过镁离子结合了三个绝对保守的天冬氨酸残基。细菌 RNA 聚合酶的活性位点已经从很多高分辨率结构中得到了研究。因为 RNA 聚合酶是同源的，所以关于 DNA、RNA 和核苷酸的结合，与催化作用一样，更进一步的讨论将转移到对真核生物 RNA 聚合酶 pol Ⅱ 的描述中。

7.2.2 σ 因子

几乎所有的 σ 因子（σ factor）都属同一个同源家族，即 σ^{70} 家族。它们完全服务于转录这一生命过程，在细菌细胞中是持家（housekeeping）基因[a]。大肠杆菌含有 6 个不同的 σ^{70} 因子，*B. subtilis* 含有 18 个，而 *S. coelicolor* 含有 63 个。这些不同的 σ 因子对不同的 DNA 序列具有特异性。如果完全靠它自身，σ 因子没法结合到启动子上去，但是可以结合到 RNA 聚合酶上，同时发生了一个有意义的构象变化，然后识别 DNA 的 − 35 位及 − 10 位的片段，最后使得核苷酸 − 12 与 +2 之间的双链被解旋，RNA 聚合酶可以开始转录基因。在转录的启动阶段之后，σ 因子从 RNAP 的核心上解离下来。当然也有抗 σ 因子的蛋白质存在，它可以特异地结合到 σ 因子上，抑制它们结合到启动子上去。

σ 因子的水解蛋白片段以及结合到 RNAP 核心上的完整蛋白的结构均已被确定（图 7.6）。这个因子具有一个延伸的“U”形结构。N 端区域是非常柔性的，还没有结构被解出。蛋白质的主要部分是包含了 4 个由螺旋结构搭建的结构域，它们由可被蛋白酶降解的柔性连接链连接。这些结构域被命名为 ND1、ND2、LD 和 CD，它们包含一个保守序列的区域。一些保守序列涉及 σ 因子结合到 RNA 聚合酶的核心位置。这些柔性的连接链在行使功能时允许有大的构象改变。

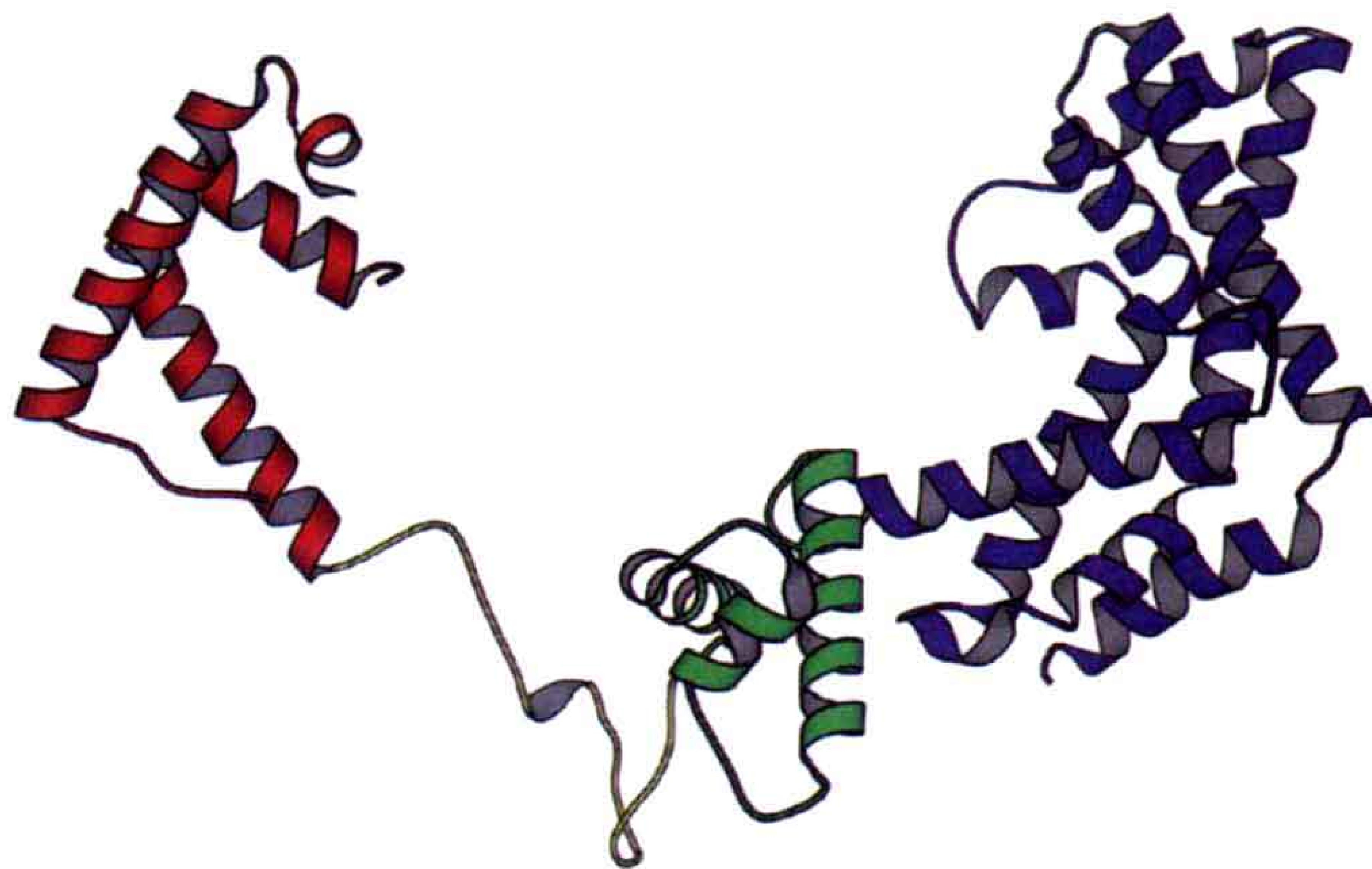


图 7.6　当 *T. thermophilus* 中 σ^{70} 结合到了 RNAP 核上时的结构。这一结构缺少了 N 端无序的 73 个氨基酸残基。这一结构完全是由 α 螺旋和柔性环组成。这一结构域已经被命名为 N 端结构域 ND1（也被称为 σ_2），含有 8 个 α 螺旋（蓝色）；而 ND2 (σ_3)，含有 3 个螺旋（绿色）。连接结构域（LD，黄色）没有二级结构，C 端结构域（σ_4）含有 4 个螺旋（红色）。σ_2 与 -10 区域（TATAAT）有结合力，σ_4 与 -35 位的共有序列 TTGACA 区域有结合力。一个螺旋 - 转角 - 螺旋（helix-turn-helix）模式在 σ_4 中出现，这与在很多转录因子中找到的结构很相似（见 7.6.3 节），其可以结合到 DNA 的这一区域。

σ 因子结合到钳子结构外侧的 β′ 亚基上。ND1 结构域与 RNAP 的核心部分具有很大的接触面。“U”形分子折叠到了 β′ 亚基的一部分中，U 结构域的外侧表面与 β′ 亚基 N 端不保守的部分相互作用 (图 7.5)。

---

a 在所有细胞中都会被表达的基因。

7.3 真核生物转录

在真核生物中存在三种 RNA 聚合酶，pol Ⅰ、pol Ⅱ 和 pol Ⅲ（图 7.7）。pol Ⅰ 只转录包含两种大核糖体 RNA 的基因。所有的编码蛋白质的基因都由 pol Ⅱ 来转录。pol Ⅲ 转录小 RNA，如 tRNA 和核糖体 5S RNA。对于 pol Ⅰ，它含有 14 个亚基，现在有一个分辨率在 12 Å 的冷冻电镜结构。pol Ⅲ 含有 17 个亚基。细菌 RNA 聚合酶核心部分的亚基（$\alpha_2\beta\beta'\omega$）是与所有真核生物 RNA 聚合酶的亚基同源的（表 7.1）。

表 7.1 不同 RNA 聚合酶之间的保守亚基的关系

RNAP	通常的保守亚基					不保守的[b]
细菌[a]	β′	β	α[Ⅰ]	α[Ⅱ]	ω	—
古细菌	A′+ A″	B	D	L	K	7
真核生物的 Ⅰ	Rpa1	Rpa2	Rpc5	Rpc9	Rpb6	9
真核生物的 Ⅱ	Rpb1	Rpb2	Rpb3	Rpb11	Rpb6	7
真核生物的 Ⅲ	Rpc1	Rpc2	Rpc5	Rpc9	Rpb6	10

a 细菌 α 亚基（α[Ⅰ]和 α[Ⅱ]）是完全相同的，但是与古细菌和真核生物的聚合酶中相应的是不同的蛋白质。
b 在真核生物中，一些附加的亚基都是比较常见的和保守的。

RNA 聚合酶 Ⅱ（pol Ⅱ）含有 12 个亚基，但是两个亚基（Rbp4/7）形成一个异质二聚体可以从核上解离下来。pol Ⅱ 是三个 RNA 聚合酶中最受调控的，存在几百个激活子和阻遏子。这些基因特异的转录因子结合到了 DNA 的序列片段上（调控的或者强化的因素），可以控制的基因达到 50 kb。除此之外，一般的转录因子都结合到 DNA 的很接近基因的启动子区域的控制片段上（图 7.1）。它们参与了识别启动子、去阻遏，以及结合 RNA 聚合酶到启动子上等过程，还有一些附加的功能。这些一般性的转录因子（见 7.5 节）被称为 TF Ⅱ A、B、D、E、F、G/J、H 和 I（TF Ⅱ 识别了一个作为转录因子的分子到 RNA 聚合酶 Ⅱ 上）。这些转录因子中对于启动必不可少的是 TF Ⅱ B、D、E、F 和 H。一个非常重要的蛋白质是 TATA 框的结合蛋白（TBP），它是 TF Ⅱ D 的一部分。TBP 结合到启动子上导致了多亚基启动前复合物的形成。TBP 部分具有与细菌 σ 因子一样的功能。

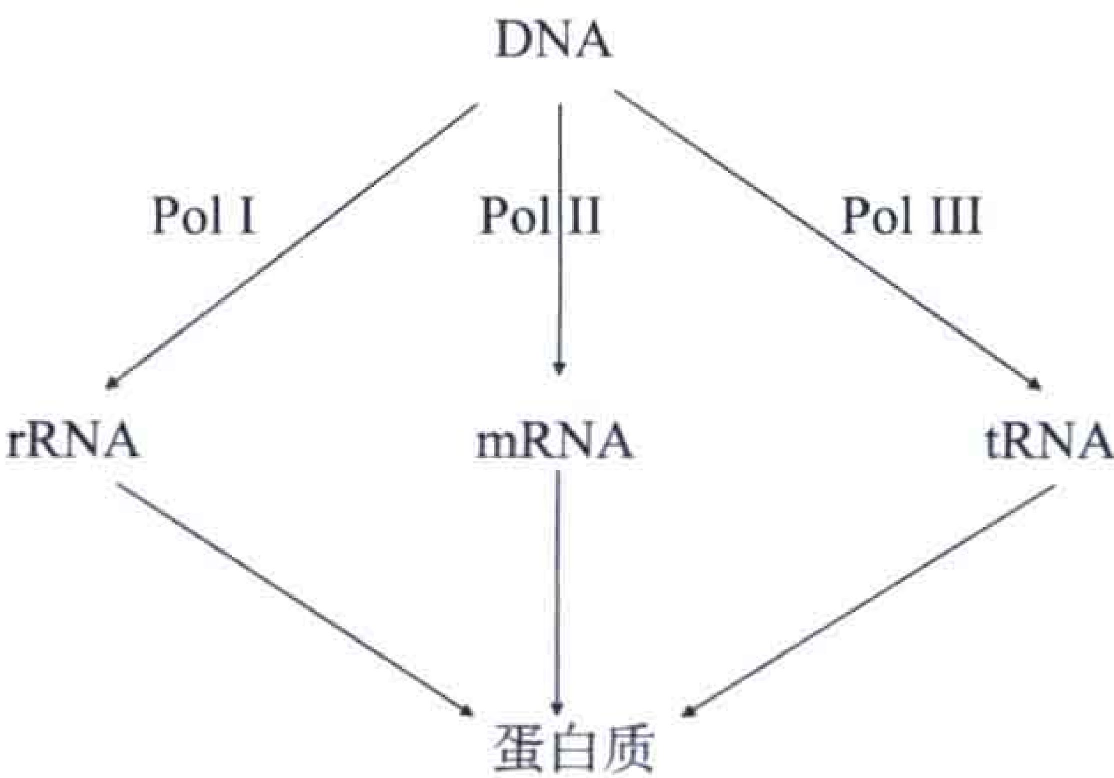


图 7.7 关于真核生物中三个不同 RNA 聚合酶功能的简单图示。

在转录以后，真核生物的 mRNA 将会被“带帽”和“加尾”，而且内含子必须被去掉。加工过的 mRNA 会被从细胞核中转运到细胞质中。mRNA 分子的寿命具有显著的变化，这也是重要的调控审核过程。

7.3.1 RNA 聚合酶 II 的结构

真核生物 RNA 聚合酶 II（pol II）包含了 12 个亚基，它的分子质量为 500kDa 左右。对于一个酵母的酶结构解析是非常困难的，Kornberg 和他的合作者用了将近 20 年的时间，才得到了一个好的结构。在这一过程中，最开始的时候是由一个 pol II 的二维晶体得到的酶的结构。这个最初的晶体结构含有 10 个核心亚基，但是缺少了 2 个不重要的残基 Rpb4/7（图 7.8）。细菌和酵母的 RNA 聚合酶的结构显示，有 5 个亚基出现在所有的生物体中，这是这一酶的最基本结构。在所有的生物体中，从活性位点的镁离子出发，大约 40 Å 半径范围内，酶的核心部分非常相似。而在边缘部分的亚基的序列和结构就会有一些不同。因此与细菌中的酶相比，这是密切相关的。

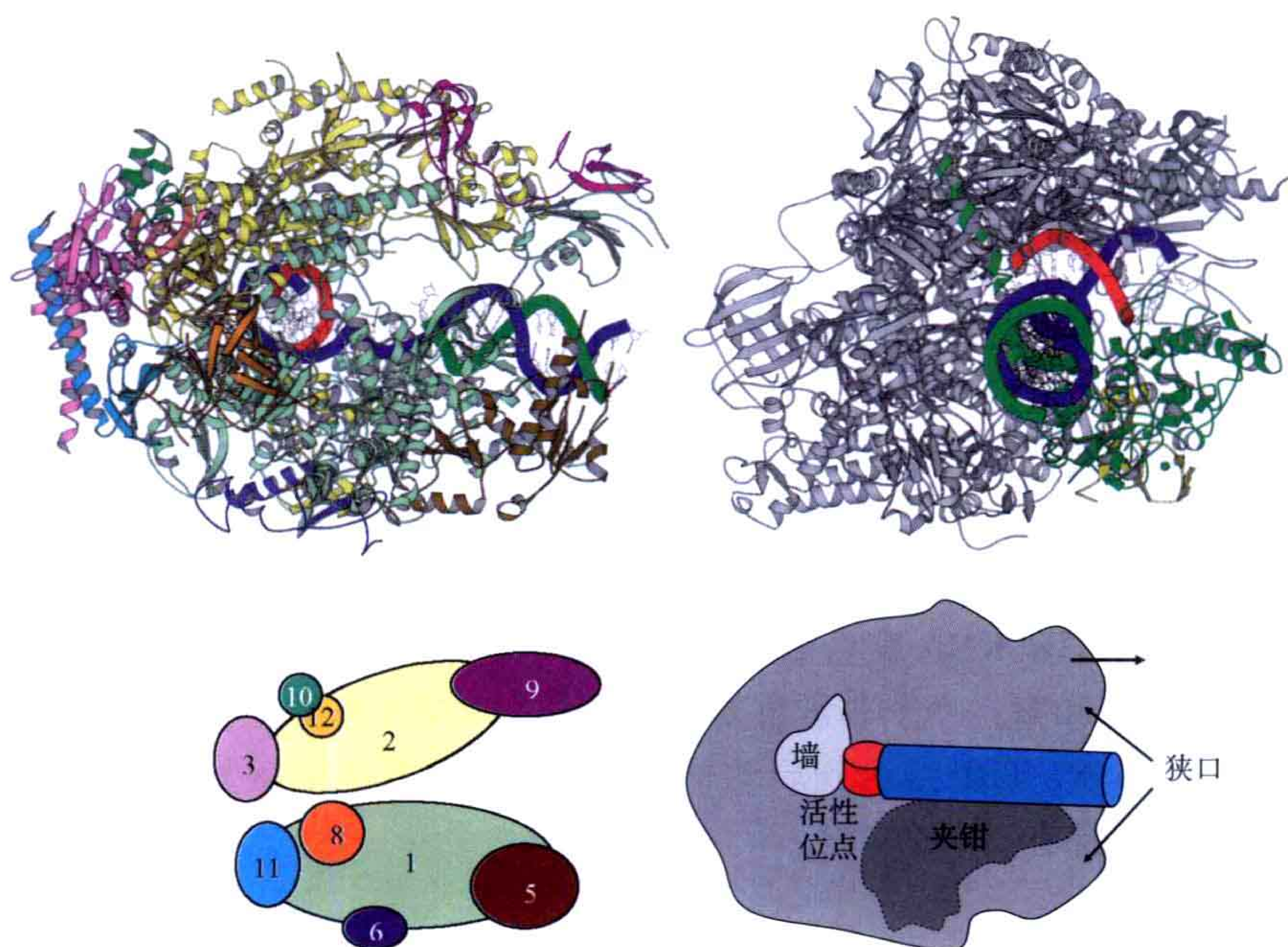


图 7.8 酵母 RNA 聚合酶 II 核心部分的结构在两个方向上的显示。左上：所有的 10 个亚基分别根据左下的简单示意图被上颜色。这里的亚基颜色与图 7.5 中的细菌酶的亚基颜色是根据同源性一一对应的。右上：显示了从上往下看的下游 DNA。桥螺旋和钳结构（the bridge helix 和 the clamp）被标记成了绿色（PDB：2E2H）。右下：pol II 中心位置特征的简单阐释图。DNA 的结合位点在右侧，与狭口和夹钳（the jaw 和 the clamp）相连接。活性位点在分子的中间部位。因为被称为墙（the wall）的这一结构，这里有一段短 DNA-RNA 杂合体被排列在了垂直的方向上。RNA 聚合酶沿着双链 DNA 向右移动。

酵母 RNAP 中最大的两个亚基，即 Rbp1 和 Rbp2，分别对应于细菌 RNAP 中的 β′ 亚基和 β 亚基（表 7.1）。它们形成了两个狭口结构，分别在裂缝的两侧。亚基 3、10 和 11 连接固定了这两个狭口结构。这些亚基就相当于细菌酶中的 α 亚基二聚体，即异质二聚体 Rpb3-Rpb11，成了亚基 Rpb1 和 Rpb2 的自组装的核，与细菌中的情况是一样的。

很多亚基都是通过附加的 β 模式实现相互作用的，也就是一个亚基的片段延伸到邻近亚基的片层中。Rbp12 是 Rbp2 和 Rbp3 之间的桥梁，参与到了两个附加的 β 模式中。pol Ⅱ的表面几乎完全带负电荷，除了活性位点裂缝处之外，那里排列的是带有正电荷的残基。Rpb1 和 Rpb2 的折叠看上去是唯一的，它们通过不同的方式相互作用。在活性位点周围，它们联合给出了一个单折叠。Rpb1 的一条螺旋桥接连了狭口之间的裂缝（这就是桥连螺旋）。Rpb2 的 C 端延伸到了相反的一侧。

表 7.2 给出了关于聚合酶结构细节的总结。酶核心部位包含了 Rpb1 和 Rpb2 的区域，形成了活性中心，涉及组装的亚基有：Rpb3、Rpb10、Rpb11 和 Rpb12。这里有三个模型是关于酶催化机理的，其是在酶的不同状态下，相对于核心位置改变了它们所处的位置。它们倾向于 DNA 结合裂缝的一侧。上面的狭口是由 Rpb1 和 Rpb9 部分形成的，以及以 Rpb2 为裂片（lobe）。下面的狭口大都含有 Rpb1 和 Rpb5。夹钳（the clamp）在裂缝的一侧，主要含有 Rpb1 的 N 端区域和 Rpb2 的 C 端区域。夹钳由三个锌离子结合区域来稳定。总的来说，8 个锌离子在 pol Ⅱ中被找到。夹钳中的一个环，被称为“船舵（the rudder）”，参与了 RNA 从和 DNA 的杂合体上解离下来的过程。

表 7.2　RNA 聚合酶的结构模块和结构域及它们的功能

名称	真核生物中的亚基	位置	功能
上颌（upper jaw）	Rpb1, Rpb2, Rpb9		
下颌（lower jaw）	Rpb1, Rpb5		
颌叶（jaw-lobe）	Rpb1, Rpb9, Rpb2 的侧翼部分	上颌	
搁板（shelf）	Rpb1, Rpb5, Rpb6	下颌	
突起（dock）	Rpb1		与 TF Ⅱ B 相互作用
墙（wall）	Rpb2	活性位点的后壁	使得 DNA 模板链发生一个 90° 的方向转变，并且形成 DNA-RNA 杂合链，以便催化反应
夹钳（clamp）	Rpb1N, Rpb6, Rpb1C, Rpb2C		夹钳是酶的移动部分，与双螺旋 DNA 和 DNA-RNA 杂合链相互作用
船舵（rudder）	Rpb1	夹钳的环	分裂 DNA-RNA 杂合链
盖子（lid）	Rpb1	夹钳的环	分裂 DNA-RNA 杂合链
拉链（zipper）	Rpb1	夹钳的环	两条 DNA 链可以在拉链后边重新结合
叉环 1（fork loop 1）	Rpb2		分裂 DNA-RNA 杂合链
叉环 2（fork loop 2）	Rpb2		阻止 DNA 的非模板链在持续过程中开始形成转录泡
桥螺旋（bridge helix）	Rpb1		在活性位点。在 +1 位置接受模板碱基
扳机环（trigger loop）	Rpb1		在活性位点的柔性环处，如果底物结合是正确的话，可以经过一个构象变化来形成螺旋对。扳机螺旋的形成减小了入口开口的宽度

7.4 RNA 聚合酶的活性位点

不管是细菌中，还是真核生物中的 RNA 聚合酶在不同底物及不同酶催化状态的结构都已经被研究过了。所有的多聚酶的催化机制看上去都是非常接近的，因为它们具有高度的保守性。但是研究还没有给出一个完全连贯的机制解释图像。这些信息一开始都来源于 *T. thermophilus* 的 RNAP 和酵母 pol Ⅱ。

RNA 聚合酶的活性位点可以由镁离子的位点来鉴定。这一金属一般在晶体结构测定中被其他的金属取代了。规则的 B- 型双螺旋 DNA 结合在 Rpb1 和 Rpb2（在细菌中是 β 和 β′) 之间的裂缝中，与位于活性位点的金属离子相遇。大概有 20 个碱基对可以容纳在这个裂缝之间。裂缝中的 DNA 是下游的 DNA，是可以被转录的部分。在起始阶段，启动子没法结合到这个裂缝中，但是可以结合到酶的其他部分。在真核生物的 pol Ⅱ 中，除了 Rpb1 和 Rpb2， Rpb5 也有可能还有 Rpb9 与 DNA 相互作用。Rpb9 属于上颌，就像 Rpb1 一样，包含有两个锌指（zinc-finger）结构，这两个锌指是由一条链连接起来的，在 Rpb1 的 β 片层（β-sheet）中形成了一个外部的 β 折叠（β-strand）。夹钳结构形成了裂缝的一边，并且与距活性位点的 DNA 非常远的下游区域相互作用（图 7.8）。这个夹钳从一个没有结合 DNA 的开放结构到结合 DNA 的闭合结构，或经过一个旋转 30 Å 的很显著的构象变化。

在转录中，DNA 模板会解旋形成转录泡（transcription bubble）。在这个泡中，转录出的 RNA 与 DNA 的模板链形成一个 8~9 个碱基对的杂合链（图 7.4）。这个杂合链不能沿着双链 DNA 线性延伸，因为 Rpb2 部分会阻止这一过程。这一部分被称为墙（wall）。RNA-DNA 杂合链恰恰相反，会在垂直的方向上延伸（图 7.9）。远离杂合链区域，DNA 和 RNA 链就会分裂，而两条 DNA 链会重新结合。

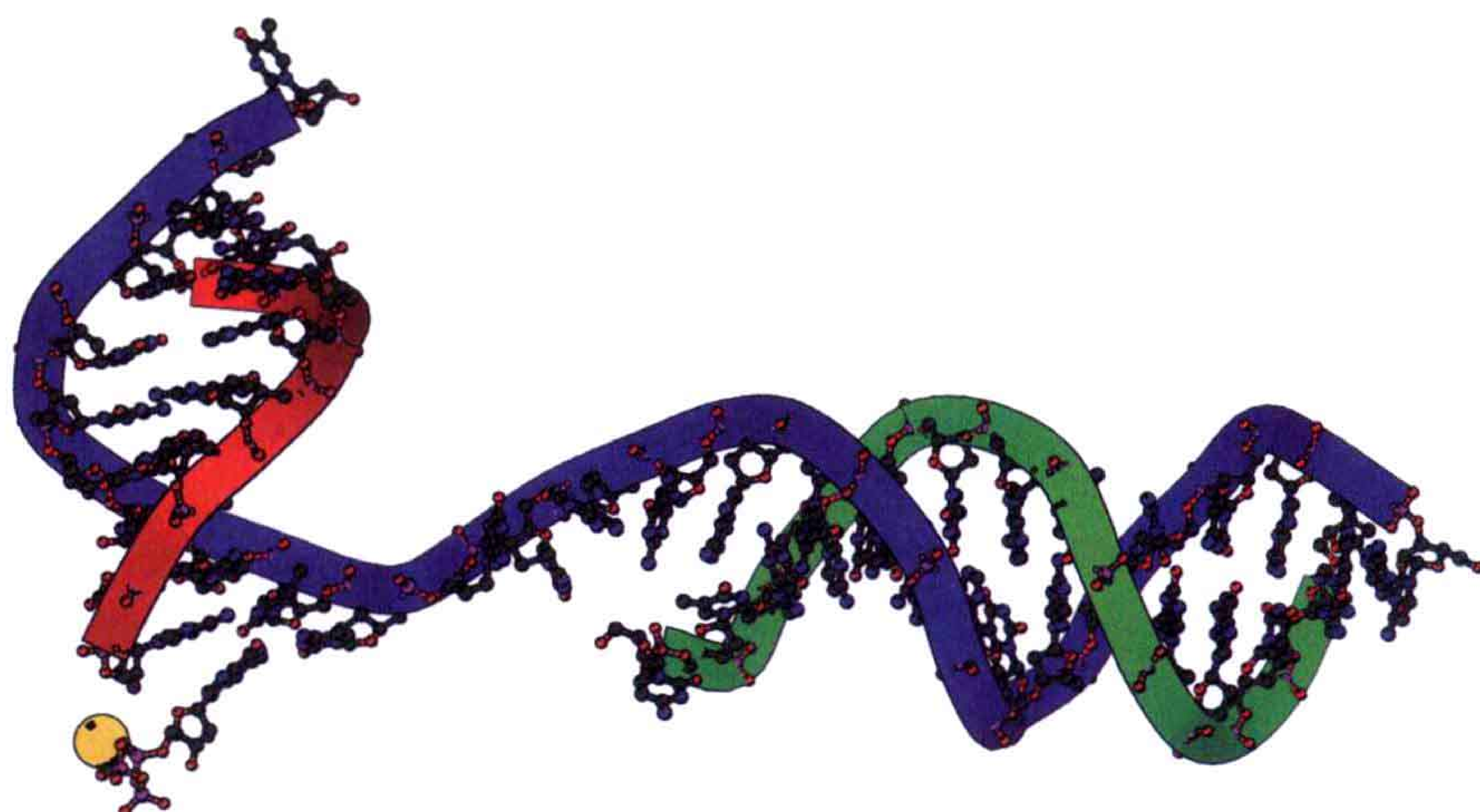


图 7.9 DNA 和 RNA 结合到 RNA 聚合酶 Ⅱ 的延伸复合物的构象。起催化作用的镁离子用黄色球来表示（PDB：2E2H）。

模板链DNA的转录泡开始于DNA的+1位置，其刚好是DNA-RNA杂合子形成的活性位点。在一些条件下，在–2和+3位置的碱基对也被观察到是转录泡的一部分。DNA碱基对的方向是在+1位置转变了90°，形成了杂合螺旋的开始（图7.11）。这个方向在一定程度上是因为酶中的螺旋，其被称为桥联螺旋。转录泡的下游边沿也是由于酶的特征而被称为Rpb2的叉环1和叉环2（图7.10）。

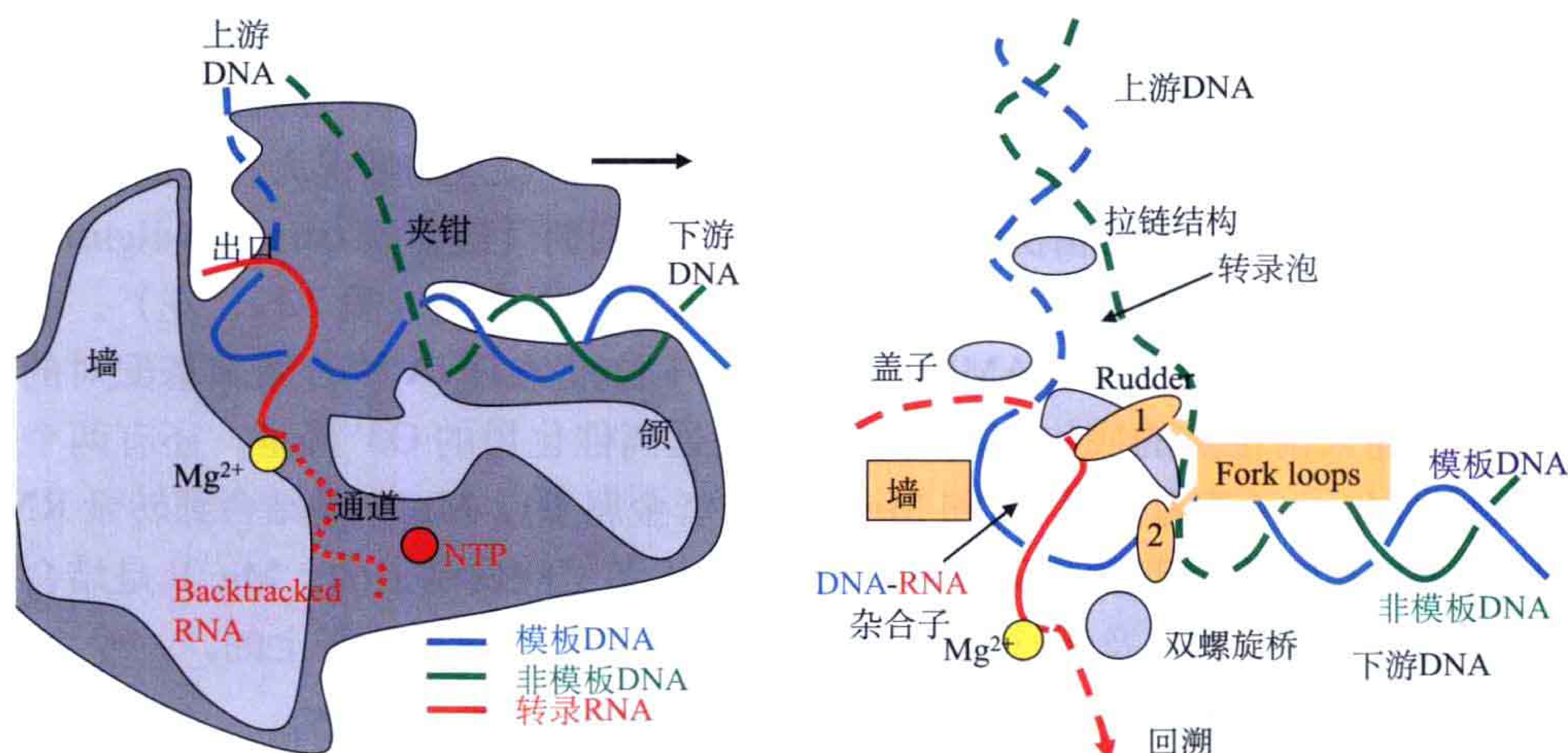


图7.10　左：pol Ⅱ的功能性组成部分的全局图。可以看到颌在下右部分，而夹钳是在双螺旋DNA的上边。杂合子DNA-RNA由上游的墙（wall）稳定。镁离子（黄色）在这里相当于金属A和Mg Ⅰ。核苷酸（NTP）通过下边的通道进入活性位点。右：酶引导核酸的一些特征。叉环2使得双螺旋DNA解旋成转录泡，这样模板链就能进入活性位点中，但是非模板链进入了另一个通路。船舵的作用是使得杂合螺旋被分裂成RNA链和DNA链，这样可以使得双螺旋DNA，在拉链结构之后重新形成。

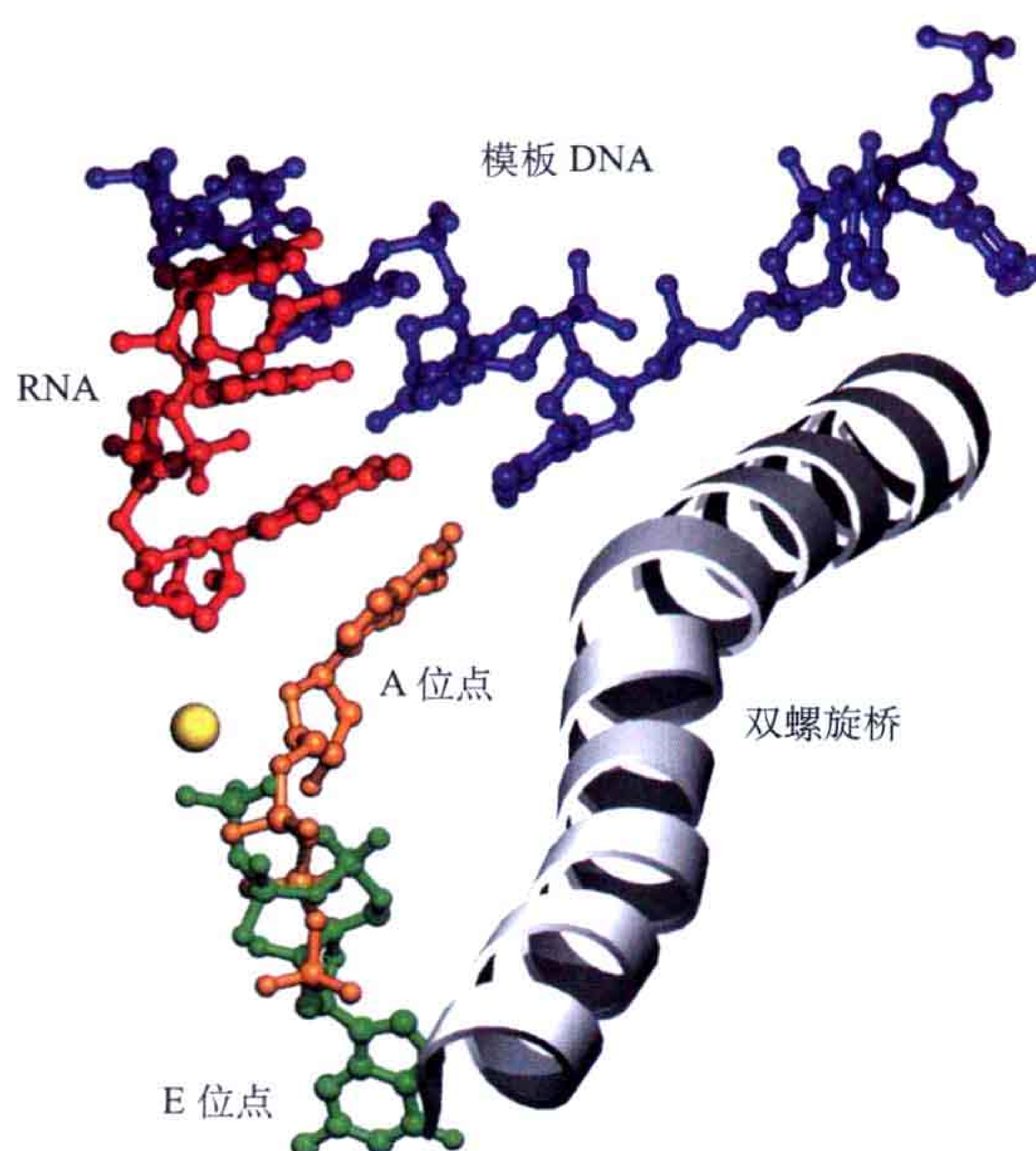


图7.11　引入核苷酸的A位点和E位点。只有同源的核苷酸可以结合到A位点上。在这种情况下，2′脱氧GTP可以结合到这两种位点上，但是大部分是结合在E位点上（PDB：2E2I）。

在转录中，两个重要的 RNA 聚合酶的状态是转录前和转录后的状态。在第一个状态时，核苷酸只是加到了 RNA 转录产物上，依然占据增加的或者“A”位点。在转录后的状态时，这一位点是空的，准备好了与新的 NTP 结合从而在 +1 位置与 DNA 核苷酸配对。

在酵母的 pol Ⅱ 中，为进入的核苷酸准备的两个结合位点已经被确定，即 A 位点和一个“E”位点。所有的核苷酸可以结合到 E 位点上，但是只有那些与模板完全配对的序列才可以结合到 A 位点。A 位点的三磷酸的位置与 E 位点的磷酸的位置非常吻合，但是 E 位点的糖环和碱基都指向远离模板的方向，这与 A 位点的情况大相径庭。

在细菌体系中，有不同的情况被发现。当一个抑制剂［链菌素 (streptolydigin)］出现在体系中时，除了插入位点以外，还有一个插入前位点存在（图 7.12，左）。在任何一种情况下，进入的核苷酸（AMPcPP）是与 +1 位置的 DNA 核苷酸碱基配对的。尽管如此，在插入前位点的底物核酸的 α 磷酸是远离催化用的 O3′ 羟基。还有两个镁离子，被称为 Mg Ⅰ（或者金属 A) 和 Mg Ⅱ（或者金属 B)。Mg Ⅰ 是结合到转录 RNA 上的最后两个核苷酸桥连磷酸上及 Rpb1 环保守的天冬氨酸残基上的。Mg Ⅱ 是结合到插入前位点的核苷酸三磷酸及任意的 Rpb1 和 Rpb2 中的天冬氨酸残基上的。

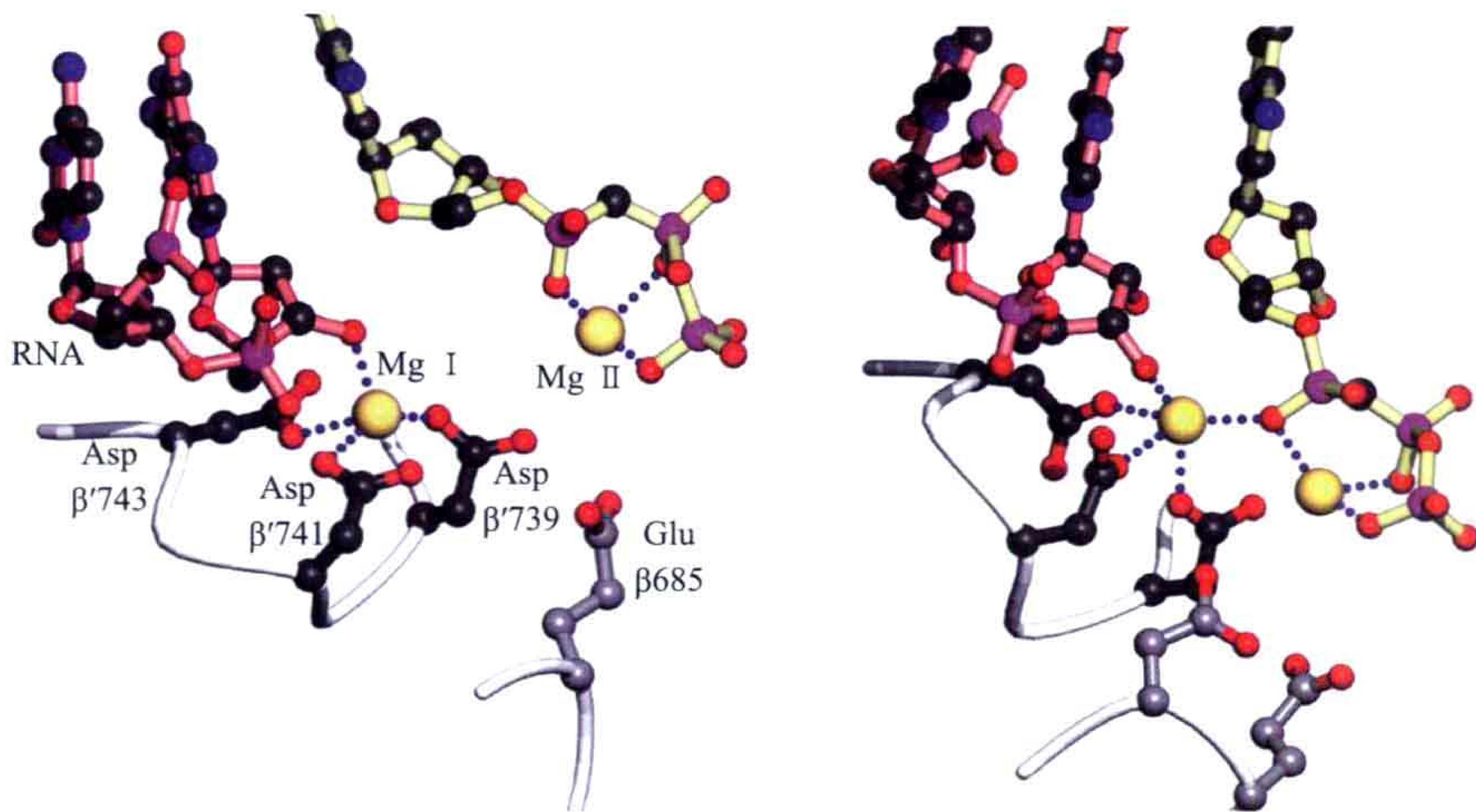


图 7.12　左：在含有抑制子链霉素 *T. thermophilus* RNAP 的晶体结构中被观察到的插入前位点中的核苷酸。进入的核苷酸（在这种情况下，AMPcPP 含有黄色的键）在 +1 位置与 DNA 模板链碱基互补配对，但是被包含的 α 磷酸与增长的 RNA 链距离太远（PDB：2PPB）。右：核苷酸 (AMPcPP) 结合到了 *T. thermophilus* RNAP 在没有抑制剂的情况下的插入位点，两个镁离子此时靠很近。Mg Ⅱ 与一个水分子配位（没有显示）。一个桥联的水分子对于两个金属缺少的配位点来说或许是合适的（没有显示）。这个水分子可以是催化或者水解焦磷酸解离中的一个参与物（PDB：2O5J）。

当抑制剂被去掉时，将看到一个不同的结构（图 7.12，右）。这里的 AMPcPP 分子正好处在插入位点。焦磷酸盐可以由 NTP 水解得到，而核苷酸被包含在增长的 mRNA 中。

T. thermophilus 的酶将经过一个有趣的构象变化，与核苷酸结合到插入前和插入位点相吻合。这个桥联螺旋是一个活性位点基本的组成部分，在转录后状态的空的 A 位

点上，它支撑了模板链在一侧的 +1 位置。当结合了 AMPcPP，扳机环（trigger loop，TL）会折叠成两个螺旋，并且延伸这个侧面的 TL 成为两个长的反平行的螺旋，这可以被称为三螺旋（图 7.13）。其连同桥联螺旋成为三螺旋泡。三螺旋残基决定了进入活性位点的入口的大小。这一非常闭合的构象是 RNAP 的活性构象。利链霉素阻止了扳机环的构象转移，这就是酶的活性构象。

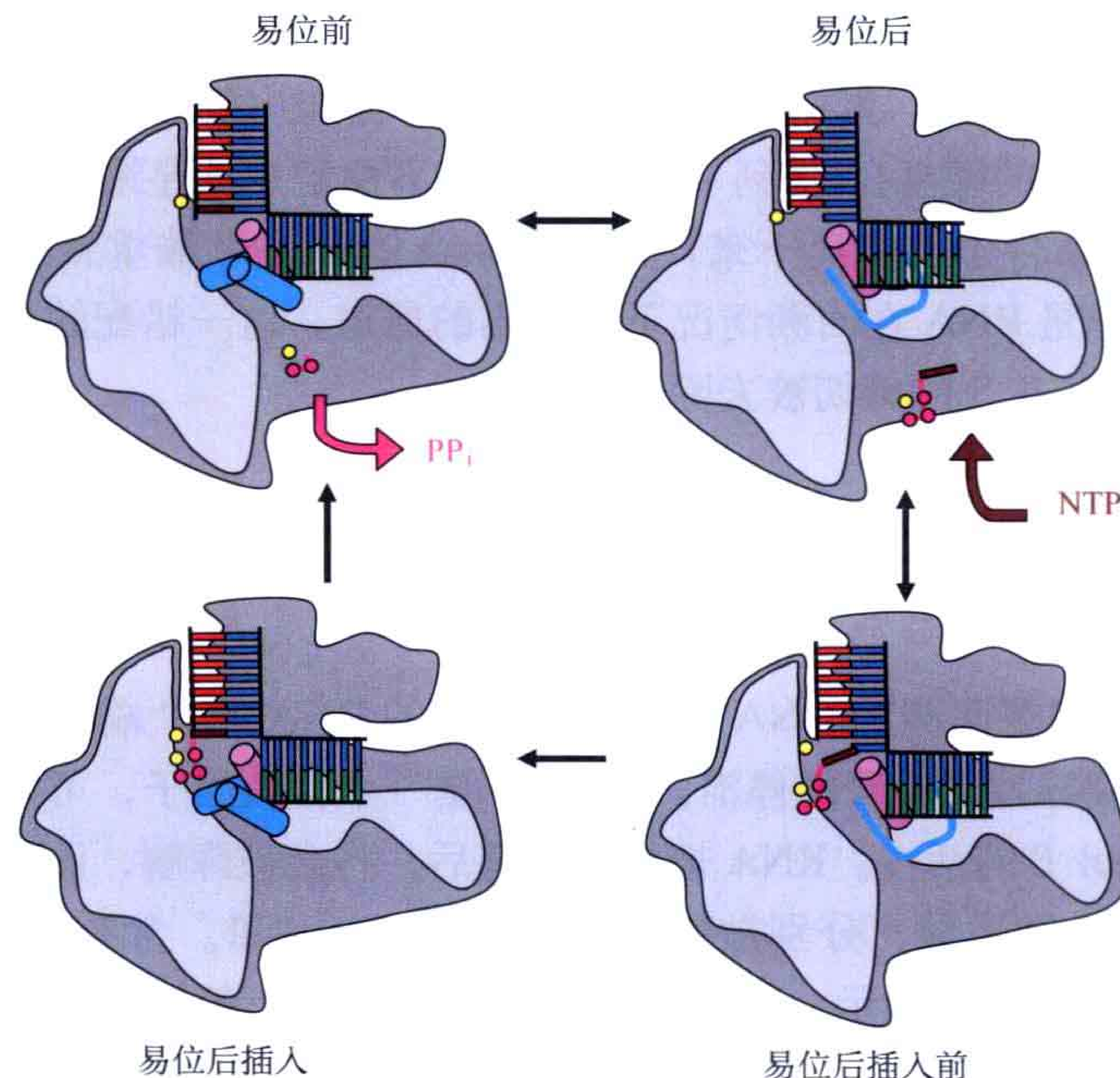


图 7.13 转录中 RNA 聚合物延伸的一个循环。图中显示了 pol Ⅱ的轮廓以及模板 DNA 链（蓝色）和非模板的 DNA（绿色）与 RNA（红色）的杂合体。桥螺旋（紫色）和扳机环（亮蓝色）也被显示了。金属 A 和 Mg Ⅱ用黄色表示。上右：一个 NTP 分子（表示成一个褐色碱基和三个紫色的磷酸）和一个镁离子（黄色）通过通道被吸收，在活性位点处，与 DNA 模板链的 +1 位置碱基配对。下右：在含有抑制剂（利链霉素）的情况下，NTP 结合在平面插入的位置。下左：NTP 被包含到了“A”位点。扳机环（TL）改变了它的构象使得周围的螺旋都延伸成为扳机螺旋（TH）。上左：RNAP 的催化态。一个同源的核糖核酸被结合到增长的 RNA 链中。焦磷酸盐基团和 Mg Ⅱ解离。随后的步骤是转移，使得接下去的 DNA 核苷酸进入活性位点而被转录。

7.4.1 RNA-DNA 杂合螺旋和回溯

当转录达到 10 个残基的长度时，RNA 将会从 DNA 的模板链上解离下来，而这时两条 DNA 链也会重新结合。三个从夹钳中衍生的被称为“船舵”、“盖子”和“拉链”（“rudder”，“lid”和“zipper”）的环分别行使了杂合子解离中的功能（图 7.10）。其中盖子不是特别的保守，与杂合子的相互作用也不是很强，它扮演了一个 RNA-DNA 链解离的空间角色。与杂合子远离的 RNA 结构是一个单链，但是还有堆积能。RNA 链穿过了酶的表面。

在转录过程中，RNA 聚合酶在向前（聚合）和向后移动（回溯）之间不断反复。扳机环的突变不但提高了向前的转移，还增强了回溯的过程。回溯在启动阶段被应用，主要是在 DNA 被损坏时进行校对。转录在起始阶段不断的重复合成短 RNA[未成熟的尝试 (abortive attempt)] 直到最少 10 个核苷酸被合成后才开始。在这个长度以下，RNA 将会从杂合子上降解。当转录达到了 20 个核苷酸的尺寸时，它就变得足够稳定了。

核糖核酸很可能比脱氧核糖核酸更加具有特异性，因为其可以同时识别核糖糖链和 DNA-RNA 杂合子螺旋。蛋白质是与杂合螺旋非标准构象互补性很高的。一个核糖核酸或者一个不正确的碱基在 –1 到 –5 区域时，会不稳定并且导致回溯。在这种情况下，RNA 的 3′ 端将从杂合子上解离下来，而这个杂合体会在 5′ 端重新形成。NTP 进入活性位点的通道同样是 RNA 在回溯情况下被伸出的通道。这一错配的核苷酸将会通过活性位点转录因子 TF Ⅱ S 的剪切被去除。

7.4.2 TF Ⅱ S，一个营救因子

当 pol Ⅱ遇到需要回溯的 DNA 序列时，生成的 RNA 的 3′ 端将会被挤进活性位点下面的通道。回溯导致了转录的停滞。TF Ⅱ S 是一个营救因子，可以增加较弱的固有的 RNA 聚合酶 pol Ⅱ的活性。RNA 被挤进通道后，将会被降解，而转录将重新开始。

TF Ⅱ S 有三个结构域，分别被称为结构域Ⅰ、Ⅱ和Ⅲ。含有 pol Ⅱ的 TF Ⅱ S 结构域Ⅱ和Ⅲ复合物的结构显示，结构域Ⅱ结合到了含有一条长连接链的下颌，这条链伸到通道和小孔中（图 7.14）。不管是被挤出的 RNA 还是 TF Ⅱ S 的结构域Ⅲ都与小孔非常的吻合。结构域Ⅲ是一个锌指结构域，其一个 β 发夹环形成了一个 pol Ⅱ 活性中心。这样的 pol Ⅱ活性中心可能利用同样的活性中心执行聚合和降解两种功能。

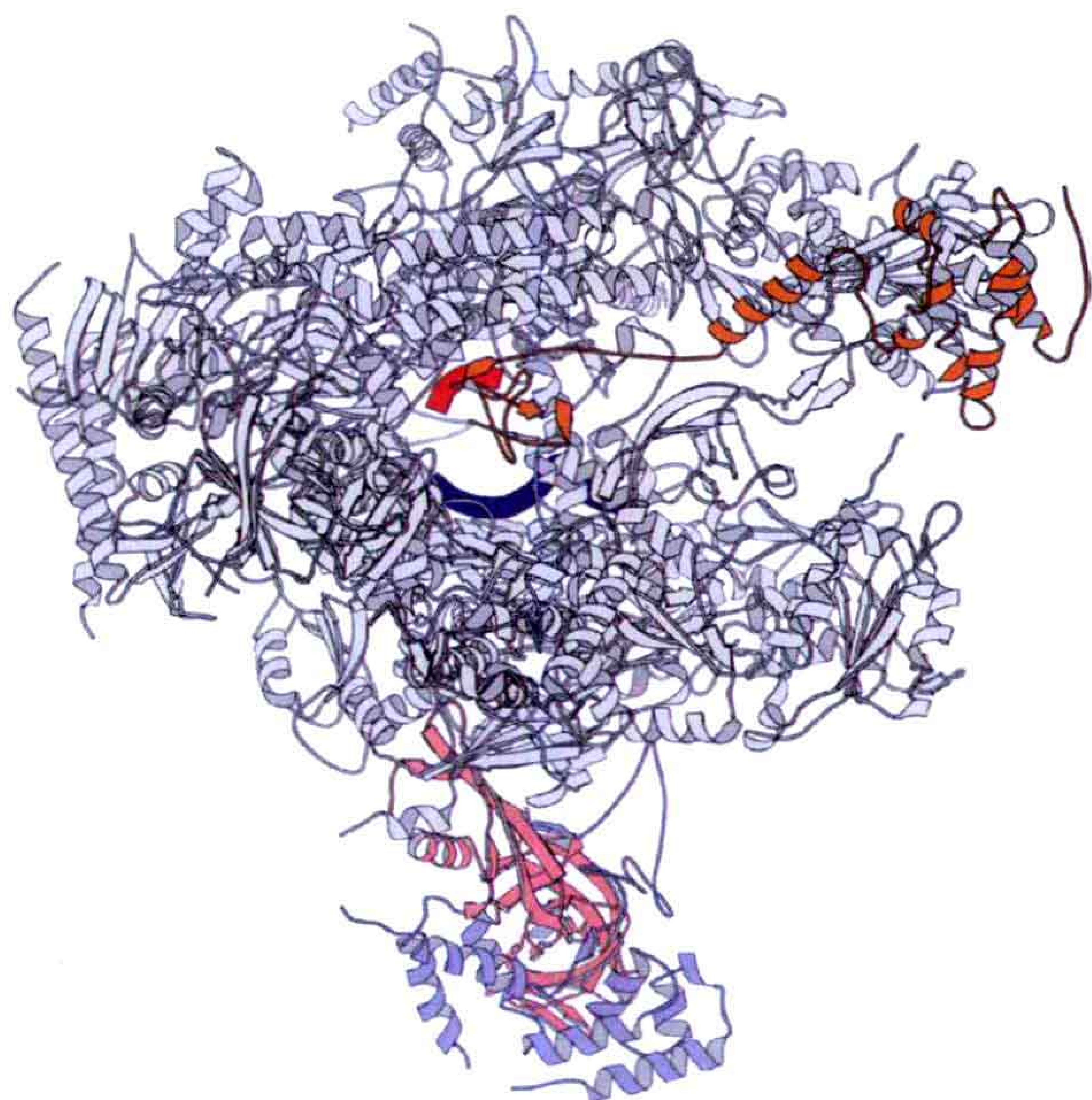


图 7.14 pol Ⅱ 含有 TF Ⅱ S 的复合物（橙色）。一个长的连接链连接了 TF Ⅱ S 的结构域Ⅱ和Ⅲ。结构域Ⅱ结合到上颌和结构域Ⅲ，部分地阻止了进入活性位点的通道（PDB：1Y1Y）。这一复合物包含了 Rpb4（蓝色）和 Rpb7（粉色）。

7.5 真核生物中一般的转录因子

7.5.1 转录的起始

RNA 聚合酶Ⅱ是没法识别启动子的，它需要基因转录因子的帮助，才能产生 mRNA。这些因子对于转录的起始来说是最先需要的。大部分的转录因子是多亚基的复合物。表 7.3 列出了这些转录因子中的一部分及它们的尺寸。一个起始前复合物（preinitiation complex, PIC）是由 polⅡ、TBP、TFⅡB 及 TFⅡF 在 TATA 框处形成的，复合物中除了包含蛋白质和 DNA 之间的相互作用外，还有蛋白质与蛋白质之间的相互作用。

表 7.3 RNA 聚合酶Ⅱ中的转录因子

名称	亚基的数量	总分子质量 /kDa	功能	结构细节
TFⅡA	2~3	50	开始于一些启动子	TBP-TATA-TFⅡA
TFⅡB	1	40	PIC 结构的中心	TFⅡB-polⅡ
TFⅡD-TBP	1	30	结合 TATA 框	TBP-TATA, TBP-TATA-TFⅡB, TFB-TATA-TFⅡA
TFⅡD-TAF	13	880	协助 TBP 识别启动子	只是一些亚基的
TFⅡE	2	90	与 TFⅡF 相互作用，需要与 TFⅡH 结合	没有
TFⅡF	3	160	像一个 σ 因子。把 polⅡ带到启动子上。TFⅡE 和 TFⅡF 参与了链的分离过程	只有因子
TFⅡH	9	530	解旋酶，可以激活打开启动子 DNA。具有激酶活性	没有
中介物	20~30	1000	配合转录的激活和抑制	只是一些亚基的

转录的起始对于蛋白质的表达来说是至关重要的一步。对于真核生物转录中的关键步骤——起始的总结性表述如下。

(i) 一个基因的特异性激活子蛋白识别了染色质高级结构中基因的控制序列。它与组蛋白乙酰转移酶相互作用，此时结合基因的组蛋白八聚体开始乙酰化，并且与 DNA 的结合变得不那么紧密了。

(ii) 此时的组蛋白被移动或者移除了，因此此处的基因就可以被转录了。染色质的重构蛋白在一个基于 ATP 的过程中起作用。

(iii) TBP 在基因激活子蛋白的帮助下识别启动子。TBP 结合到了 TATA 框上，并使得 DNA 弯曲。

(iv) TFⅡD 的 TAF 与 TATA 框的弯曲 DNA 的旁侧序列相互作用。

（v） TF Ⅱ B 与 TF Ⅱ D 和 DNA 相互作用。它把启动子定位到聚合酶上，这就是启动前复合物（PIC）。

（vi） PIC 经过一个大的构象变化，使得 DNA 双链开始解离，并且把模板 DNA 链放到了 pol Ⅱ的活性位点。在这一步骤中，TF Ⅱ E、TF Ⅱ F 和 TF Ⅱ H 扮演了很重要的角色，通过引起 DNA 中的扭应变（torsional strain），从而促进链的分离。这一过程需要 ATP。

（vii） RNA 合成开始的起始步骤。

一个激活的基因习惯于启动多个 RNA 聚合的循环。一个新的 pol Ⅱ分子也可以从转录再起始中间物开始启动一个转录（图 7.15）。它包括基因特异性激活蛋白、TF Ⅱ D、TF Ⅱ A、TF Ⅱ E、TF Ⅱ H 和中介物。从另一个方面讲，TF Ⅱ B 和 TF Ⅱ F 会解离，并且必须参与每个新的启动过程。

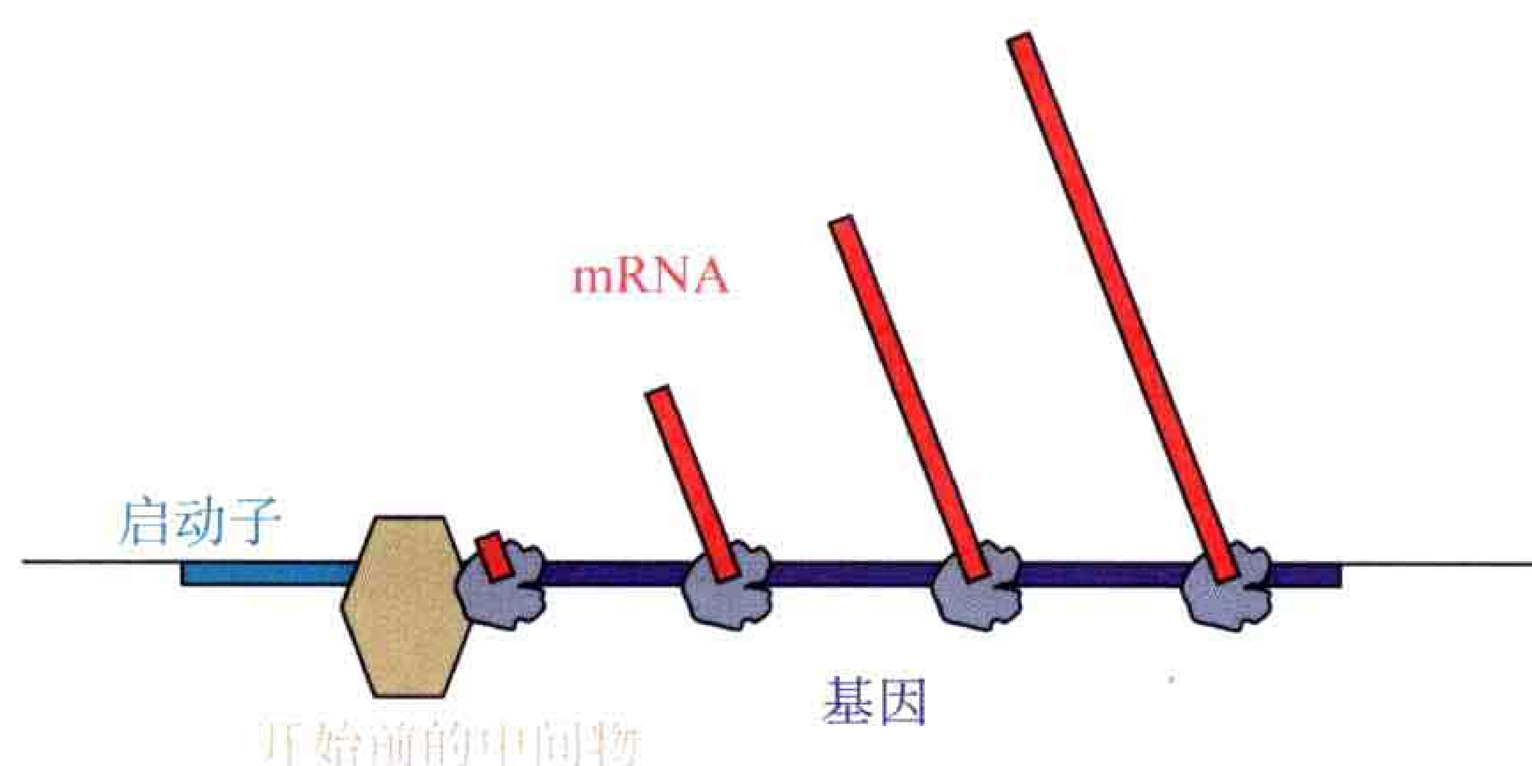


图 7.15　基因特异性激活蛋白，转录因子 TF Ⅱ D、TF Ⅱ A、TF Ⅱ E、TF Ⅱ H 和中介物形成了一个开始前的中间物，在启动子位置，一些 pol Ⅱ在 TF Ⅱ B 和 TF Ⅱ F 协助下可以开始一些转录的循环。

因为 TF Ⅱ D 和它的 TATA 框结合蛋白（TBP）在转录起始中具有中心地位，我们将首先讨论它。

7.5.2　TF Ⅱ D

TF Ⅱ D 是一个多蛋白的复合物，包含了 TATA 框结合蛋白（TBP）和 TBP 关联蛋白（TAF）。在 *S. pombe* 中，它是一个接近 900kDa 的复合物。对于启动子识别来说，TF Ⅱ D 是重要的，同时它提供了一个骨架（scaffold），使得转录机器其他部分可以组装到它的周围。在转录起始阶段，在酵母中大部分基因是依赖于 TF Ⅱ D 的。

7.5.2.1　TATA 框结合蛋白（TBP）

就像之前提到的，TATA 框是一个由 8 个碱基对组成的启动子的一部分，位于转录开始位点之前 30~100 个核苷酸的位置。一致的序列是 TATA（A/T）A（A/T）（A/G）。TATA 框结合蛋白（TBP）对于起始前复合物（PIC）的组装来说是一个中心组成部分。

TBP 是普遍存在的，三种真核生物 RNA 聚合酶都能利用它，它是一个 30kDa 左右的蛋白质分子。这个蛋白质的 DNA 结合的 C 端部分是一个马鞍的形状，是由两个 80~90 个氨基酸残基的结构域搭建的，具有非常相似的结构（图 7.16）。它骑跨在 DNA 上，反平行的 β 片层中的 8 个 β 折叠结合到了 TATA 框的小沟中，将其部分解开，并且使得它向大沟弯曲了 75° ~80° 。

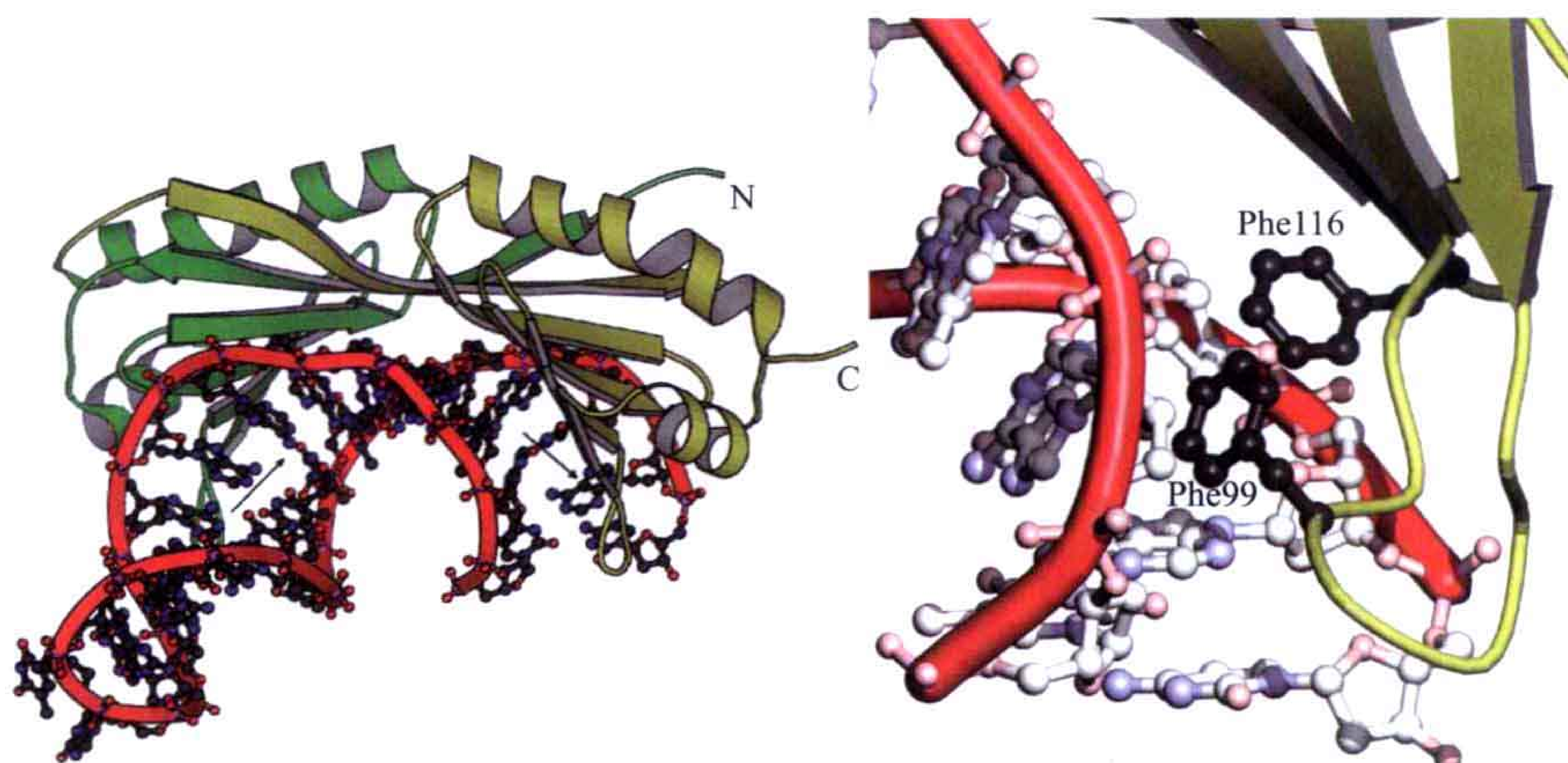


图 7.16 来自酵母的 TATA 框结合蛋白（TBP）与 DNA 发夹结构的复合物。左：这个蛋白质由具有非常相似结构的两个结构域组成（黄色和绿色）。这一蛋白质具有马鞍的形状，可以结合到双链 DNA 上。TBP 结合到 TATA 框的小沟中，并且使得 DNA 朝大沟有一个 80° 的弯曲，由箭头表示。右：在一个稍微不同的方向观察的特写，图显示了苯丙氨酸侧链插入碱基对之间引起了弯曲。有两个苯丙氨酸残基在另一个结构域中，出现了很相似的相互作用（PDB：1YTB）。

TBP 首先通过非特异性地与碱基的疏水基接触结合到 DNA 上。蛋白质经过一个很小的构象变化更加地接近 DNA：一个结构域经历了一个相对于另一个的扭转运动。一个非常重要的部分是苯丙氨酸侧链插入 DNA 的小沟中。两对苯丙氨酸的侧链插入 TATA 框两端的碱基对中，其中两个在第一个和第二个碱基对中，另外两个处在倒数第二个和最后一个碱基对之间（图 7.16）。这些插入的苯丙氨酸侧链形成了 DNA 的缠绕形状，而 TBP 剩下的部分使得 TATA 框光滑地弯曲，并且将其解开。如果撇开非特异性结合，TBP 结合是受到 TATA 框的突变影响的。除了与碱基的特异性结合，其他的一些因素看上去也是非常重要的。其可能会是 DNA 的弯曲能力，特别是对于富含 AT 序列的 DNA。

DNA 的弯曲使得本来相距较远的 DNA 部分变得更加接近了，因此其他的一些转录因子可以稳定地与上游和下游结合。TBP 的 C 端镫形物（stirrup）是在 TATA 框的上游，而 N 端的镫形物是在 TATA 框的下游。水分子在蛋白质与 DNA 之间作为媒介物使它们存在一些接触。

7.5.2.2 TBP 关联蛋白（TAF）

在酵母的缺少 TBP 的 TFⅡD 复合物中，一共有 13 种蛋白质，它们被称为 TBP 关

联蛋白（TAF）。它们对于细胞的发育非常重要，同时有助于 TFⅡD 对于操纵子的特异性结合。除了 TAF 独立的启动子以外，还有 TAF 相关的启动子存在，它涉及上游的激活序列。在酵母中，大部分的基因是与 TAF 相关的。然而，TAF 的个体功能还是存在很多不清楚的地方。

TAF 具有 9 个与组蛋白相同折叠的结构域（图 7.17）。通过 TFⅡD 对于特定的启动子结构的识别是基于这些结构域的。TAF 中的一些会被分别表达两个拷贝，形成一个异源四聚体复合物。有人提出这些蛋白质会形成像核小体一样的八聚体组装体，但是冷冻电镜结构却没有支持这一排列。

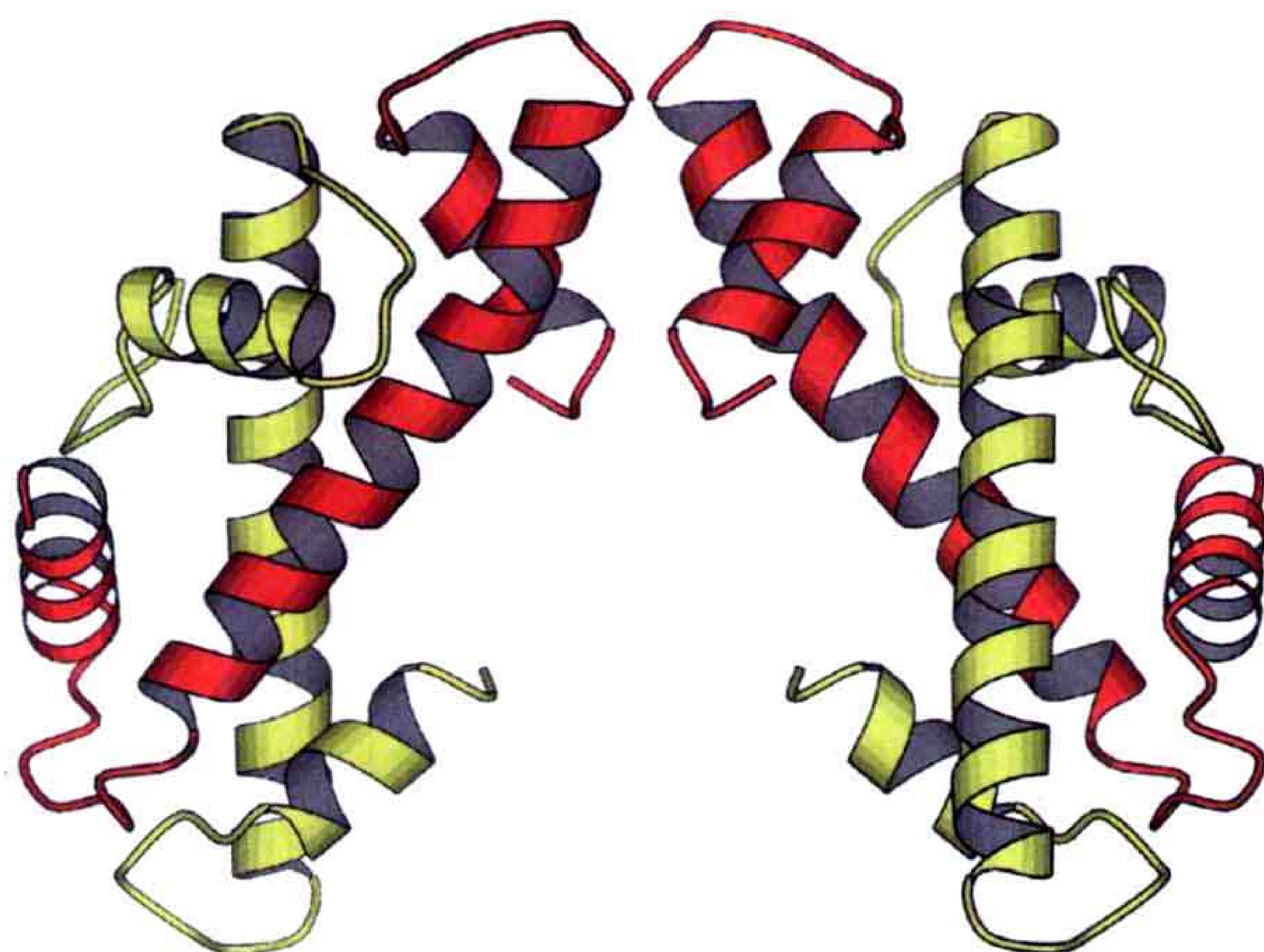


图 7.17　来自 *D. melanogaster* 的 TAF6（黄色）和 TAF9（红色）N 端结构域的两个二聚体（PDB：1TAF）。这些亚基具有与组蛋白相同的折叠类型，同时形成二聚体，但是它们不会像组蛋白在核小体中的那样通过相同的方式连接成八聚体。

TAF1 是 TAF 蛋白中最大的一个，在酵母中，其大于 1000 个氨基酸残基。在 *Drosophila melanogaster* 中，TAF1 含有 2000 个残基，并且含有两个激酶结构域，一个在 N 端区域（NTK），而另一个在 C 端区域（CTK）。酵母蛋白缺少 CTK 结构域。CTK 可以特异性地磷酸化组蛋白 H2B。除此之外，TAF1 还具有组蛋白乙酰化酶结构域和泛素活化 / 结合活性。TAF1 的 N 端区域通过结合到 TBP 的鞍（saddle）下面抑制了 TBP 与 TATA 框的结合。从另一个方面来讲，TBP 通过与 TAF 的结合和解离来实现调控 TFⅡD 的结构。

7.5.3　TFⅡB

TFⅡB 是只有一个亚基的转录因子。它的单多肽链具有一些独特的结构域，这对于形成起始前复合物（PIC）是非常重要的。它的功能是识别转录因子的起始位点，以及使得 TBP 和 DNA 结合到 pol Ⅱ 上去。连同 TFⅡB 和 pol Ⅱ，它形成了一个最小限度的起始复合物。TFⅡB 与它的同源物之间的相互作用确定了转录的方向和极性。至少对于在体外试验的转录中，它是不可缺少的。

C 端区域（TFⅡB_C）与 TBP 相互作用，并且 DNA 的 6 个或者 7 个碱基对直接与弯曲的 DNA 的 TATA 框的上游和下游相互作用。这些相互作用点被称为 B 识别序列。TFⅡB 的 N 端部分开始于一个锌带（zinc ribbon），紧接着的是一个 TFⅡB 中非常保守的部分，被称为 B 指（图 7.18）。

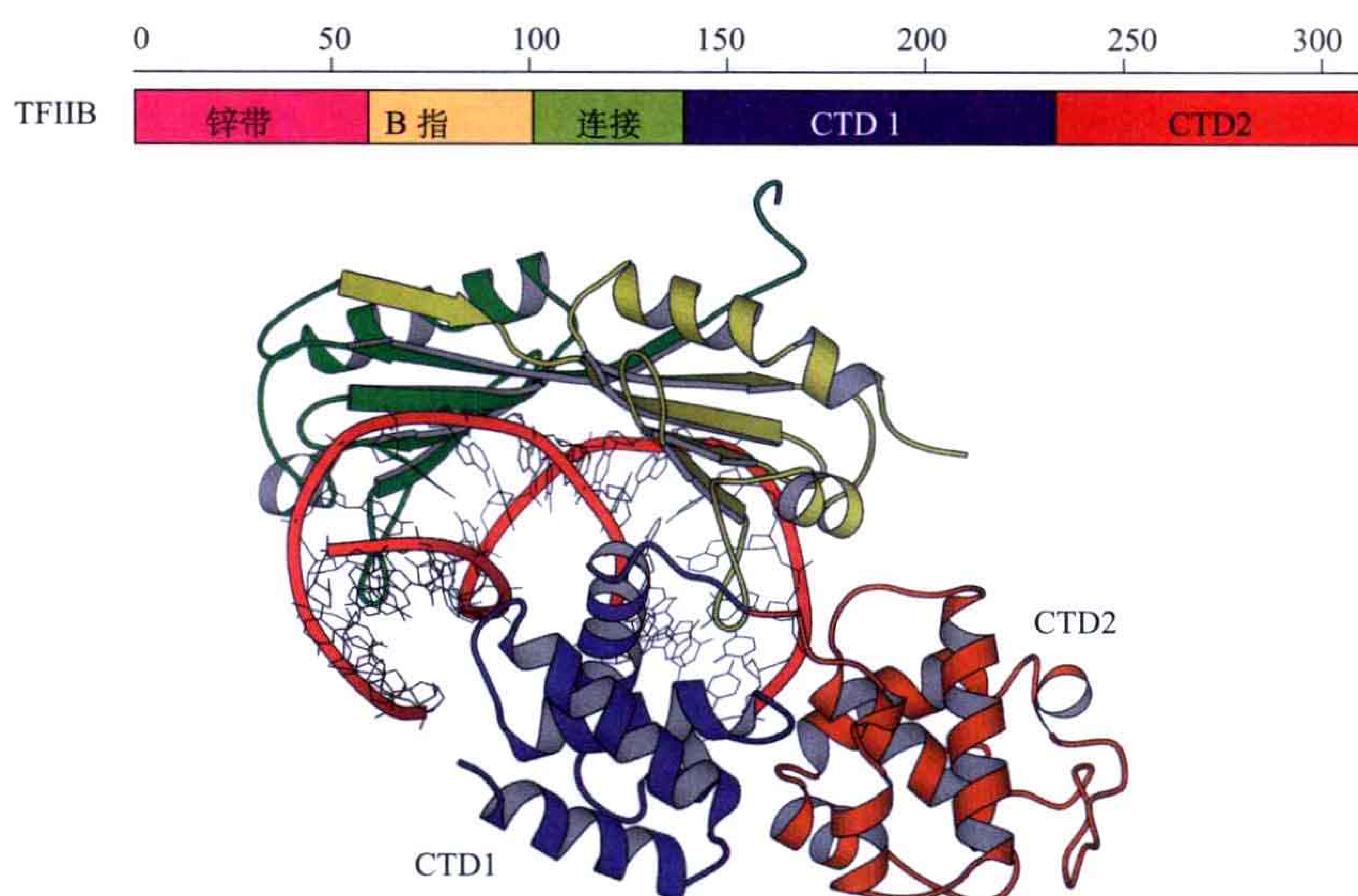


图 7.18　上：TF Ⅱ B 的单多肽链的组织图。B 指是 TF Ⅱ B 中最保守的部分。下 :TBP 和 TF Ⅱ B_C 之间的相互作用，其通过一段在 TATA 框处的 16 碱基对区域。TF Ⅱ B_C 的两个结构域分别含有 84 个残基，以及一个周期蛋白 A 的相关折叠（related fold）。TF Ⅱ B_C 结合在马鞍形状的 TBP 的 C 端镫形物上。

TF Ⅱ B_C 是由两个紧密连接的结构域组成的，这两个结构域都与细胞周期蛋白 A 相关。两个结构域的相同序列不超过 20%。在两个结构域之间，没有简单的双折叠关系，是分别由 5 个螺旋构造的。结合有 TBP 和一个 DNA 的 TATA 框区域的 TF Ⅱ B 的晶体结构显示了鞍形状的 TBP 结合到了 DNA 的小沟上，使得 DNA 朝大沟弯曲了。TF Ⅱ B_C 结合到了 TBP 的 C 端的镫形物上，并且与 DNA 的 TATA 框上游和下游编码的和非编码的链相互作用（图 7.18）。TFⅡB_C 的底面结合到了 DNA 磷酸二酯的骨架上。TF Ⅱ B_C 只是与 DNA 的大沟和小沟边沿碱基的限制性接触。这使得它们的结合是非特异性的。DNA 通过 TBP 形成的弯曲没有受到 TBP 与 TF Ⅱ B 的相互作用的影响。

N 端的锌带模式（zinc ribbon motif）结构域是通过 NMR 确定的。除此之外，还有一个与 TFⅡB 形成复合物的 pol Ⅱ结构，但其缺少了 TBP 和 DNA，尽管分辨率相当低，但其还是提供了很多更进一步的细节，电子密度图只描述了 B 指（B-finger）部分。pol Ⅱ经过一个特殊的构象改变，通过它的夹钳采取更加接近的位置结合上去了。在此结构中观察到 TFⅡB_N 的锌离子，而锌带（残基 20~54）非常适合 Rpb1 的“dock”结构域，非常接近 RNA 离开 pol Ⅱ的路径（图 7.19）。B 指（残基 55~87）穿越 pol Ⅱ的夹钳和墙进入活性位点，以同样的方式离开。

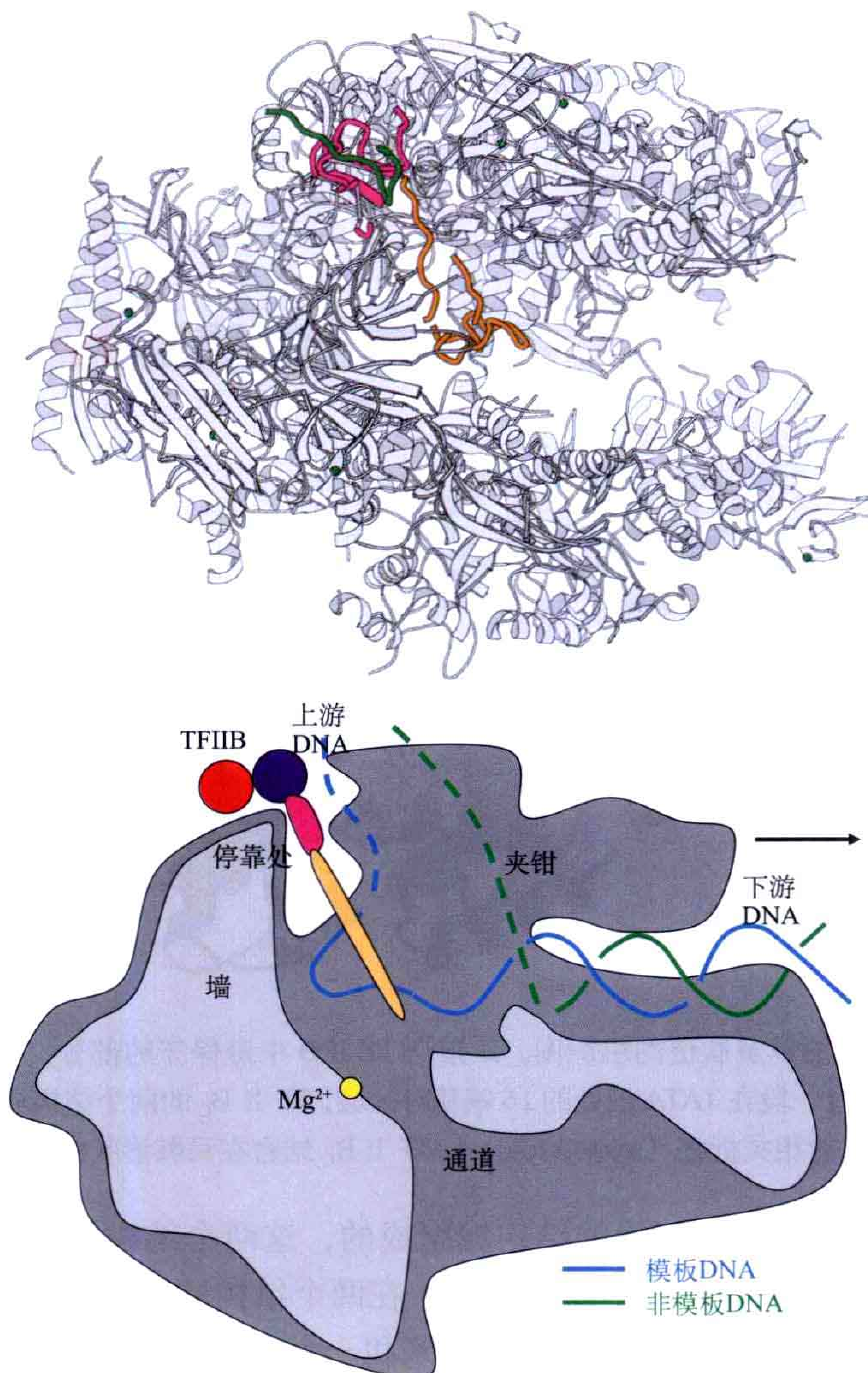


图 7.19　上：pol Ⅱ和 TF Ⅱ B 之间的复合物的晶体结构，下：在 pol Ⅱ活性位点中的 TF Ⅱ B 和它的 B 指结构的概要图。这里的各个颜色与图 7.18 中一一对应。锌带与 dock 结构域相互作用，与 RNA 的出口位点重合。B 指结构稳定了 RNA/DNA 杂合体。

TF Ⅱ B 的功能是把含有 TATA 框的 TF Ⅱ D 复合物带到 pol Ⅱ上，以这样的一种方式 DNA 可以与 pol Ⅱ 的裂缝相互作用及直接进入活性中心。这些作用通过两种不同的相互作用确定转录起始位点。这一初步瞄准（coarse setting）是由于 TF Ⅱ B_C 与 TATA 框相互作用。严格来讲，是因为 TF Ⅱ B_N 的指结构域，它与单链模板相互作用。B 指结构域稳定了结合有不少于 9 个碱基对的起始复合物。B 指结构的突变可以改变转录起始位点或者减少转录的水平。当转录的 RNA 达到了 10 个碱基的长度或者更长时，TF Ⅱ B 将会解离，这是因为空间位阻。

在 TATA 框和大约 25 个碱基对的转录起始部分之间的保守距离与执行这些工作的蛋白质的保守性有关。

7.5.4　TFⅡA

TFⅡA 不是一个主要的转录因子，因为它在一些情况下是可有可无的。然而，TFⅡA 对于特定的细胞核病毒激活物的转录刺激是被需要的。在酵母中，它是由两个亚基组成的，分别是 TOA1 和 TOA2。在人类和 *Drosophila* 中，它有三个亚基，分别为 α、β 和 γ（分子质量分别为 34kDa、19kDa 和 12 kDa)。酵母中的 TOA1 携带了哺乳动物亚基 TFⅡAα 和 TFⅡAβ 的序列。TFⅡA 与 TBP 在转录起始的第一步中相互作用，并通过防止 TBP 的二聚化来调控转录。它同样抑制阻遏子蛋白干扰 TBP 结合到 TATA 框上。除此之外，对于序列特异性转录因子，它还行使了一个共同激活子的功能。

TFⅡA 的两个重要的部分分别是一个 β 桶（β-barrel）和一个四螺旋的捆（four-helical bundle）（图 7.20）。β 桶与 TBP 的非常接近其下游镫形物区域的 N 端部分相互作用。β 桶也会在不破坏其结构的情况下与 DNA 相互作用。

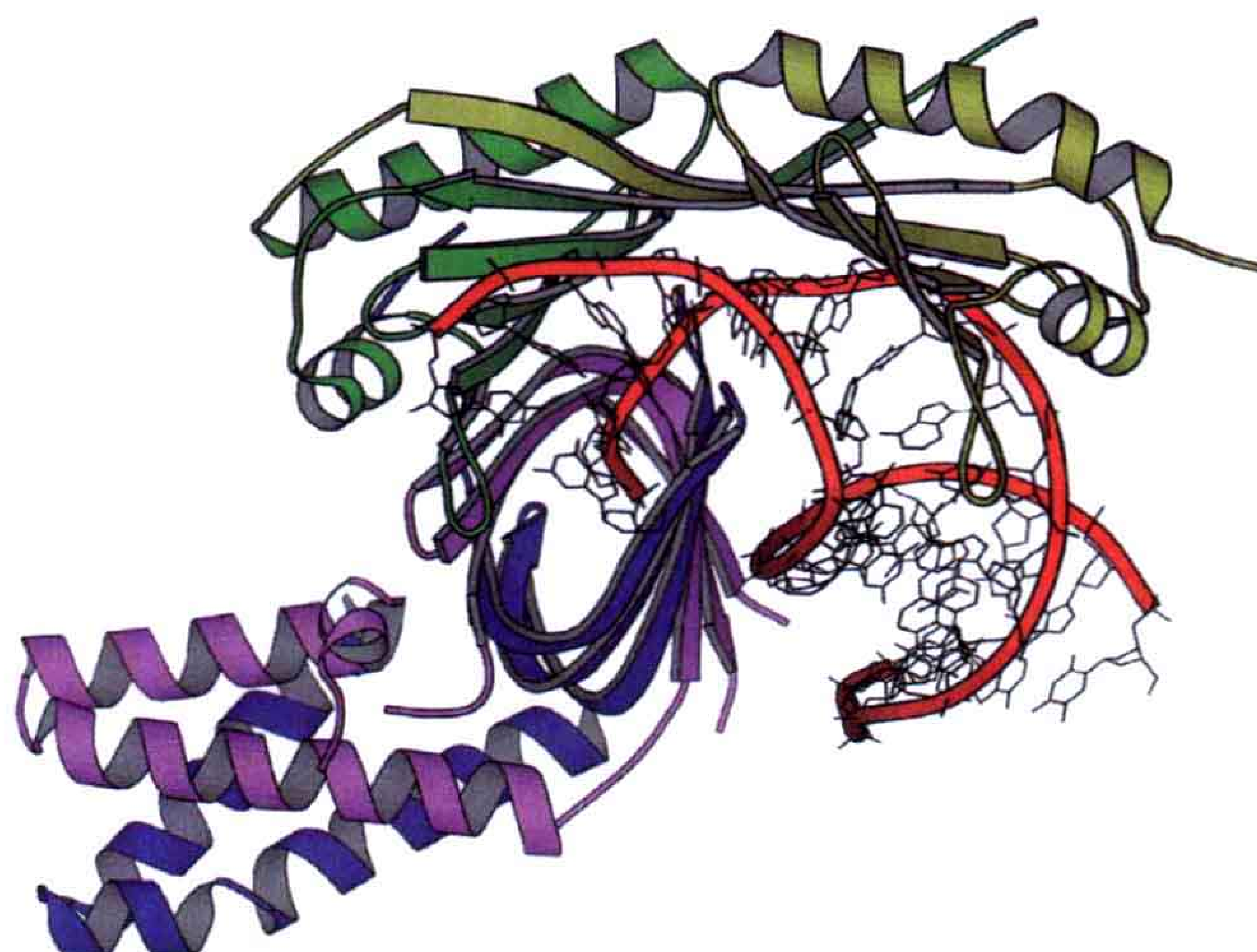


图 7.20　结合有 TBP 和酵母 TFⅡA 的 TATA 框 DNA 结构（PDB：1NH2）。TOA1（蓝色）和 TOA2（淡紫色）都含有一个 N 端、一个螺旋捆和一个 C 端 β 桶结构。TOA1 的 N 端和 C 端部分由一个 100 个氨基酸的片段连接起来，这一部分没有包含在实验中。

7.5.5　TFⅡF

TFⅡF 是由两个 RNA 聚合酶联合蛋白组成的，在人类中它们分别被称为 RAP30 和 RAP74；酵母中的分别被称为 Tfg2 和 Tfg1。Tfg2 具有与细菌转录因子 σ 明显的序列同源性。像 σ 因子一样，TFⅡF 通过 pol Ⅱ 抑制了非特异性转录起始，但是却使得聚合酶连同 TFⅡB 和 TFⅡD 都结合到启动子 DNA 上的起始前复合物。这一复合物可以结合 TFⅡE 和 TFⅡH 形成完整的起始复合物。TFⅡE 和 TFⅡF 参与了链分离的过程，并且 TFⅡF 启动了 RNA 转录中第一个磷酸二酯键的形成。

TFⅡF 的两个亚基都含有球状结构域，在终端连接有一个柔性链。这两个蛋白质的相互作用结构域的结构是非常不寻常的，它是由三个缠绕的 β 桶形成的。这两个蛋

白质的折叠具有相同的拓扑结构，而且两个亚基通过一个假想的双折叠对称轴相关。尽管如此，没有序列同源被发现。

7.5.6 TFⅡH

TFⅡH 是由 10 个亚基组成的。这些亚基可以被分成两个组，即含有 6 个亚基的中心和含有 3 个亚基的 CAK 模型。第 10 个亚基连接了这两个部分。TFⅡH 是唯一的具有酶活性的转录因子。TFⅡH 结合到 TFⅡE 上导致了启动子的打开，从而起始转录。它的激酶活性是通过细胞周期蛋白 H 激活的，使得 Rpb1 的 C 端尾被磷酸化，从而 pol Ⅱ可以从转录起始阶段进入延伸阶段。

7.5.7 媒介物

媒介物（mediator）是在 20 世纪 90 年代被发现的。在酵母中，它是一个含有 25 个亚基的复合物，全部加起来分子质量大概有 1MDa。在哺乳动物中，这可能会变成 30 个亚基左右。它的功能是配合转录激活状态或者抑制状态。媒介物支持最基本的通信，通过转录因子结合到调控序列上，或者增强 pol Ⅱ或一般的转录因子。它主要在起始和增强转录及 pol Ⅱ的 C 端结构域通过 TFⅡH 被磷酸化的过程中行使一个在基因特异性 DNA 结合转录激活子和一般的转录装置之间的界面的功能（见上节）。媒介物由 4 个特殊的模型构建：头、中间部位、尾部和周期蛋白 C（cyclin C）。前面的 3 个部分形成了核心部分，而周期蛋白 C 经常与核心部分无关联。这些模型都具有独立的功能。周期蛋白是媒介物的激酶亚基。

对媒介物结构研究的进展还不够。冷冻电镜显示了一些不同的结构形式。在相同的阶段，媒介物具有大概 400Å 的长度。在 pol Ⅱ和媒介物的复合物中，媒介物采取了这种延伸的形状。尽管如此，单独的媒介物可以折叠成更加紧密的结构。媒介物与 pol Ⅱ的最大亚基 Rpb1 的 C 端结构域具有特异性相互作用。媒介物穿过头和中间模型结合到了 pol Ⅱ的背面。

7.6 基因特异性转录因子

7.6.1 DNA 结合蛋白调控转录

基因特异性转录因子控制个体基因的转录是通过结合到 DNA 的特定位置上来完成的。在真核生物中，它们都被称为阻遏子和激活子。然而，在真核生物中，特定的转

录因子可以被用作阻遏子，也可以被用作激活子，这取决于其他的因子和修饰分子。

一些共用的结构模式已经被发现，其是完全用于转录调控的。由于 DNA 的双折叠对称，DNA 结合蛋白通常也是二聚体的。转录因子通常含有一个序列特异的 DNA 结合结构域和一个激活结构域（transactivation domain）。DNA 结合结构域在一些家族中只是一个结合到 DNA 大沟中的 α 螺旋，但是另外一些共用的模式是螺旋 - 转角 - 螺旋模式，它是在原核生物中最常见的模式，还有一些不同类型的较小的锌结合结构域，被称为锌指（Zn finger）结构（表 7.4）。

表 7.4　真核生物转录因子根据它们的 DNA 结合结构域分类的几个家族

名称	结构域的描述	举例
bZIP	亮氨酸为基础的拉链	GCN4, Fos, Jun
bHLH	亮氨酸为基础的螺旋 - 环 - 螺旋	MyoD
bHLH-ZIP		MAX, Myc
侧翼螺旋（winged helix）	变化的螺旋 - 转角 - 螺旋模式	HNF-3
同源框结构域（homeodomain）	变化的螺旋 - 转角 - 螺旋模式	POU 结构域
C_2H_2 锌指	经典的锌指	Zif268
C_4 锌指	经常是一前一后的两个一样的结构域	核受体
Zn_2Cys_6 锌指		GAL4
STAT	见第十一章	STAT1, STAT2 等

7.6.2　螺旋与 DNA 的相互作用：b ZIP 和 b HLH 转录因子

螺旋蛋白与 DNA 相互作用的最简单情形是通过一个二聚体蛋白的一对 α 螺旋，这里 α 螺旋被分开，使得它们可以正好适合去拥抱相邻的两个 DNA 大沟。含有这一简单设计的转录调控蛋白家族被称为碱性区亮氨酸拉链（basic region-leucine zipper, bZIP）。这些蛋白质的序列中都含有一个碱性结构区域，这个区域的每 7 个氨基酸残基中含有一个亮氨酸（7 个重复）。7 个周期表明一对 α 螺旋的盘绕螺旋（coiled-coil）结构。来源于酵母的蛋白质 GCN4 含有一个非常长的螺旋，这里的亮氨酸重复形成了一个二聚体相互作用（图 7.21）。在直链螺旋中一圈螺旋含有 3.6 个氨基酸残基，而在弯曲的螺旋中是 3.5 个。因此，亮氨酸残基会出现在螺旋的同一侧，每两个螺旋圈出现一次，另外，更加疏水的残基会出现在中间的位置。使用拉链这个名称是因为一开始误认为亮氨酸侧链是像拉链一样相互交错的，其实并非如此，因为它们肩并肩相互作用，形成了一个疏水核心。

这个碱性区域插入了 DNA 的大沟中。序列能够识别是因为碱性区域中的侧链可以与碱基一一相互作用。DNA 中的结合位点是一个 9 碱基对的区域，并且都具有对称性。两个 GCN4 单体结合到相同的序列上，分别在对称轴的两边。另一个关于亮氨酸拉链的例子是人类致癌基因产物 Fos 和 Jun，它们会形成一个异源二聚体。

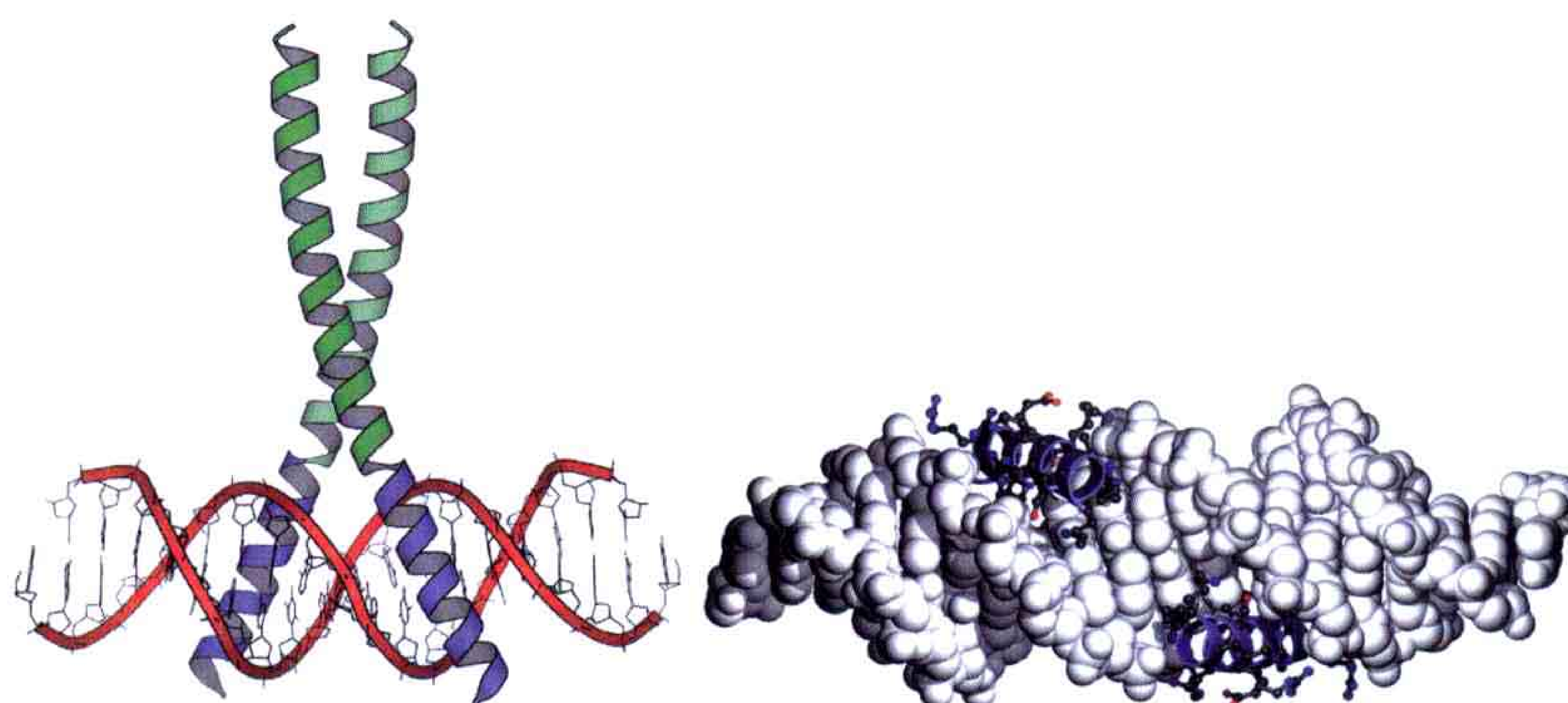


图 7.21　GCN4 的 C 端 DNA 结合域结合 DNA 形成的复合物（PDB：2DGC）。蛋白质的 N 端部分的构象还是未知的。左：结合到 DNA 上的二聚体；右：螺旋与 DNA 的大沟相匹配。

另一个二聚体转录因子家族也形成一个与 GCN4 很相似的结构从而与 DNA 结合，但不同的是它们的二聚作用。MyoD 蛋白属于这个转录因子家族。这个转录因子和一些相关蛋白涉及肝脏发育。它们可以形成同源二聚体或者异源二聚体。它们的碱性螺旋 - 环 - 螺旋（basic helix-loop-helix, bHLH）结构包含一个环区连接的两个螺旋的 C 端部分（图 7.22）。这一碱性区域使得二聚体之间通过形成四螺旋束来实现二聚。第一个螺旋的 N 端部分结合到 DNA 的大沟中。与亮氨酸拉链蛋白很相似，这一区域碱性很强，螺旋上的侧链与 DNA 大沟上的碱基对直接相互作用。

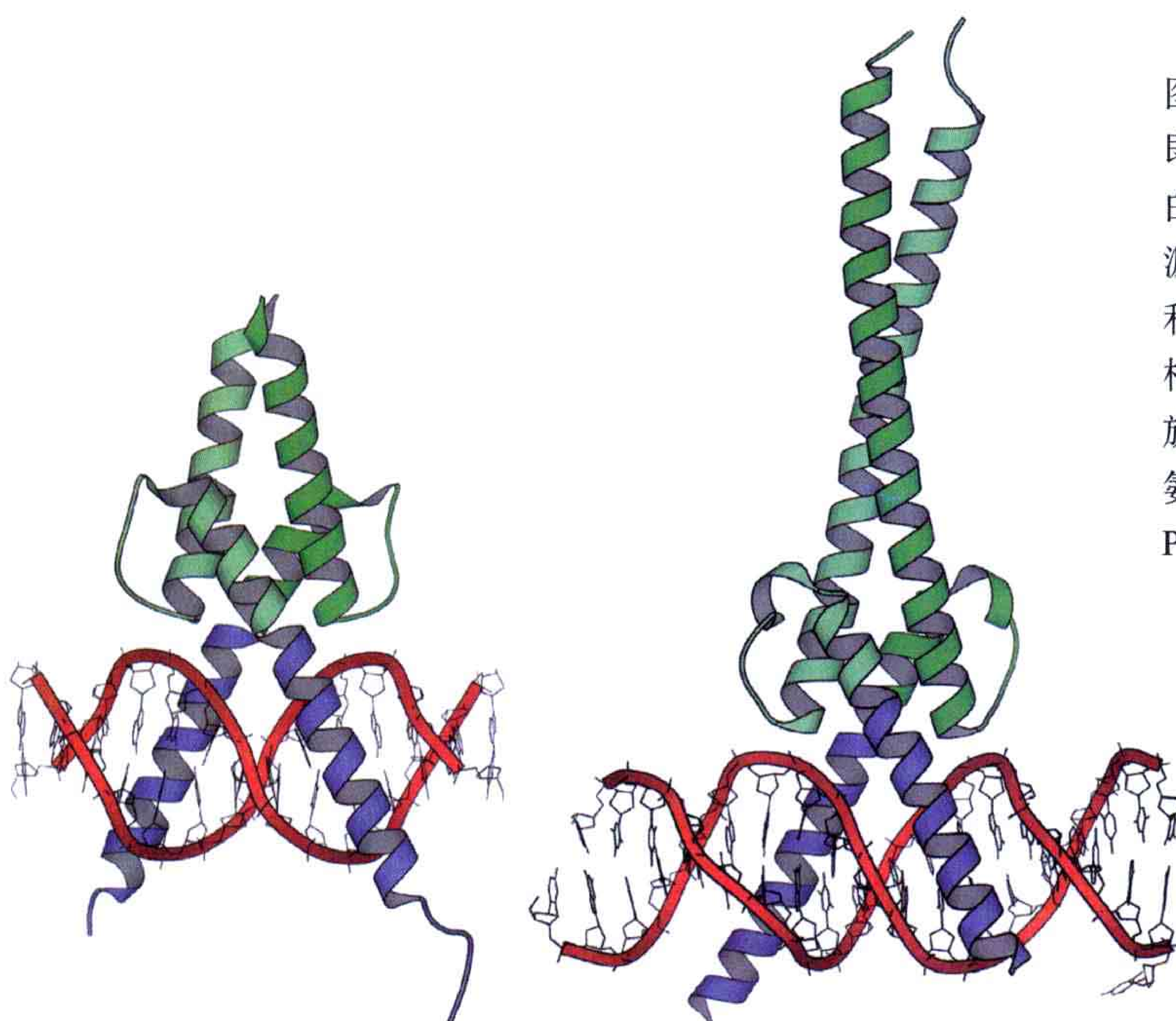


图 7.22　结合 DNA 的 MyoD，即一个碱性螺旋 - 环 - 螺旋蛋白（左，PDB：1MDY）在异源二聚体 Myc-Max 中。Myc 和 Max 单体具有非常相似的构象，也会通过一个碱性螺旋 - 环 - 螺旋区域和一个亮氨酸拉链结构二聚化（右，PDB：1NKP）。

一些致癌基因可以形成同源二聚体或者异源二聚体，它们形成的方式可以是螺旋-环-螺旋结构，也可以是亮氨酸拉链区域（bHLH-ZIP 转录因子）。其中一个例子是蛋白质 Myc，它与 bHLH 蛋白很相似，但是其模式的第二个螺旋会延伸形成另外一个亮氨酸拉链（图 7.22）。

7.6.3 转录因子中的螺旋-转角-螺旋模式

另一种常见的 DNA 相互作用类型是由螺旋-转角-螺旋（HTH）结构模式的蛋白质形成的，这种类型在细菌转录阻遏子中很常见。它们通常既可以是单体也可以是二聚体，通过螺旋结合到 DNA 的大沟中。λ 家族的噬菌体具有两个阻遏子，习惯上被称为阻遏子和 cro。这两个蛋白质都是结合到噬菌体基因组的操纵子区域的。cro 阻遏子是一个小的二聚体蛋白，大约含有 70 个氨基酸。蛋白质结合的操纵子区域是一个完全回文的、含有 14~17 个碱基对的 DNA 序列。阻遏子是一个大蛋白质，它的 N 端区域是用来与 DNA 结合的。这两个蛋白质的 DNA 结合部分都含有三个螺旋，其中的两个具有完全相对的方向，它们的螺旋轴基本是垂直的。这一排列也在很多其他 DNA 结合蛋白中被找到，这一保守的模式因此被称为螺旋-转角-螺旋模式。

cro 和阻遏子蛋白与 DNA 结合的复合物显示，螺旋-转角-螺旋模式中的第二个螺旋会结合到 DNA 的大沟中去（图 7.23）。蛋白质二聚体的对称轴与回文序列的中心相吻合，这两个单体会通过相同的区域与相邻的两个 DNA 大沟的同一侧相互作用。cro 和阻遏子蛋白的 DNA 结合部分的结构非常相似，但是二聚化后的结果是完全不同的。在 cro 中，结构域之间是通过 β 折叠相互作用的，在阻遏子中，来自两个结构域的两个螺旋之间会相互作用。

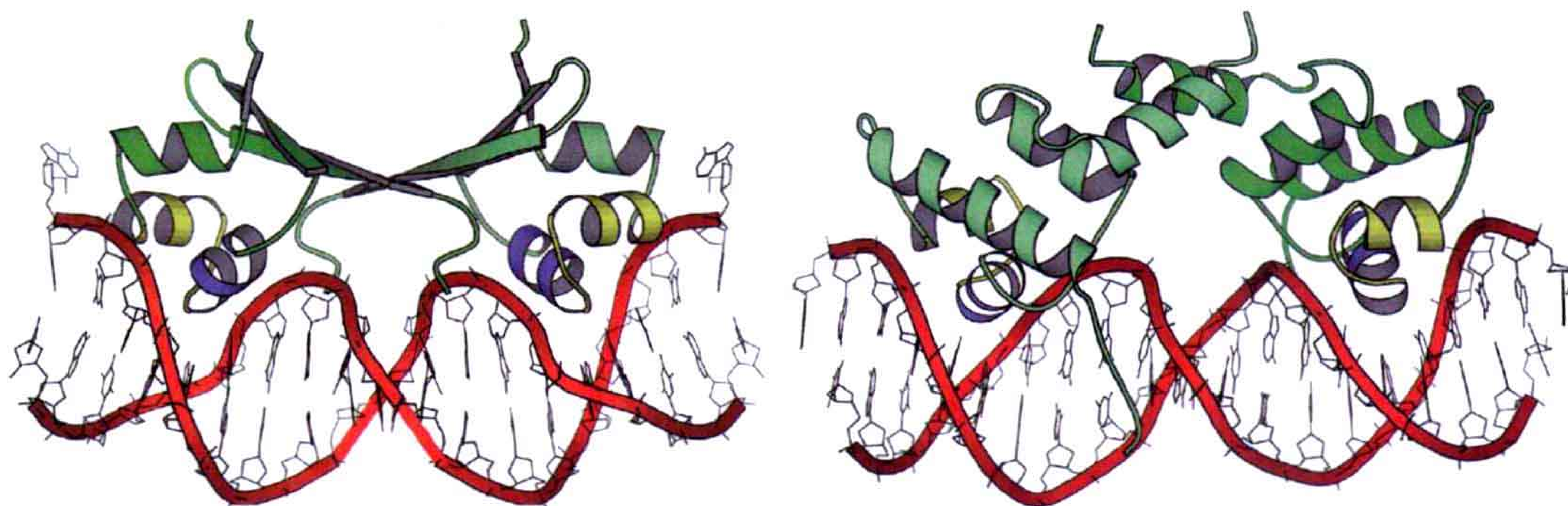


图 7.23　来自 λ 噬菌体的 cro（左，PDB：6CRO）和阻遏子（右，PDB：1LMB）二聚体结合 DNA 的复合物。阻遏子结合到 DNA 上是通过它们的螺旋-转角-螺旋模式。这个模式的 N 端螺旋被标记成黄色。识别螺旋（蓝色）结合在 DNA 的大沟上。在 cro 中，二聚体化是由来自各自单体的两个 β 折叠之间的相互作用实现的，而在阻遏子中，它们的二聚体化是因为螺旋之间的作用。

阻遏子结合的调控是转录控制的一个重要部分。在细菌中，trp 阻遏子是一个具有

相同的螺旋 - 转角 - 螺旋模式的二聚体蛋白，它的功能是调控色氨酸合成基因的转录，但是在这种情况下，它们的二聚化是不同的。

trp 阻遏子是对色氨酸浓度敏感的，它有一个结合这个氨基酸的结合位点。在没有色氨酸存在时，trp 阻遏子以不能结合到操作子位点的构象存在。当色氨酸结合上去时这个蛋白质将改变构象，使得它可以结合到操作子上，阻碍了色氨酸合成酶更多的转录（图 7.24）。

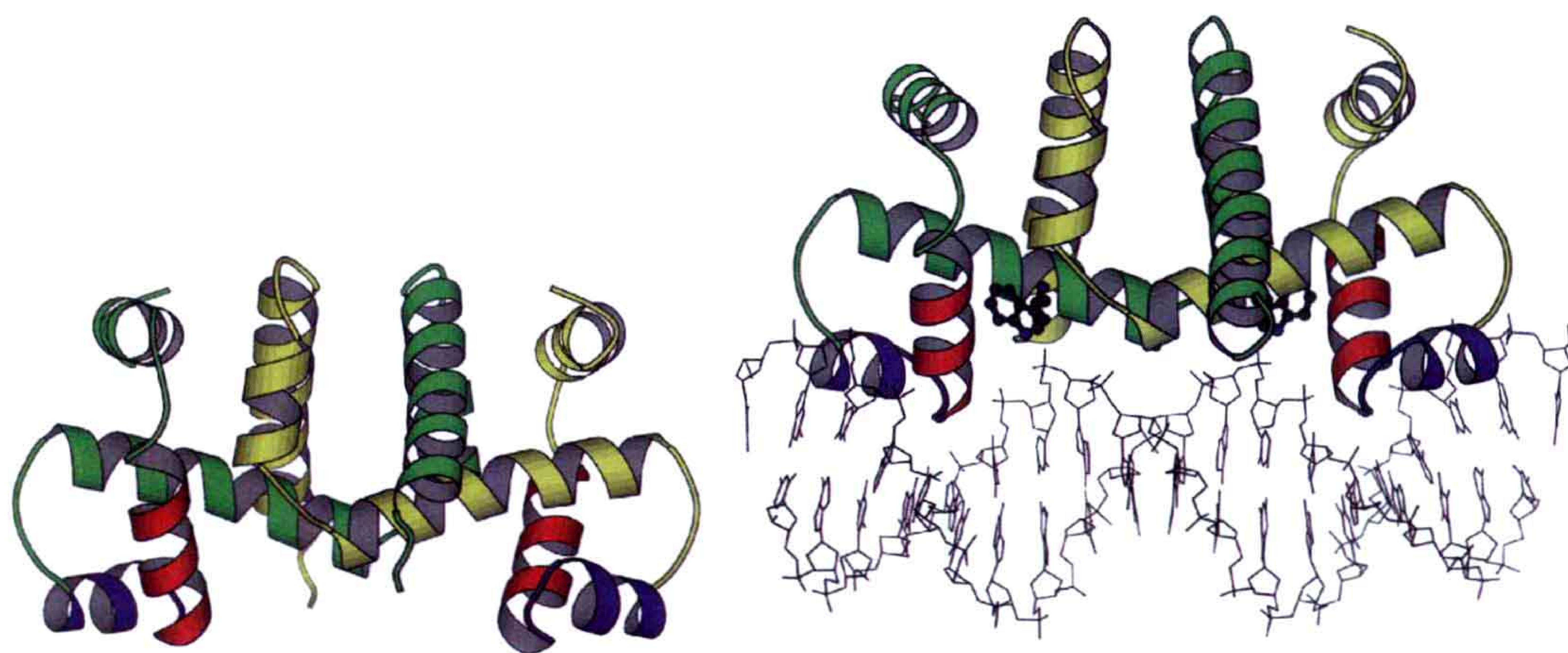


图 7.24　在大肠杆菌的 trp 阻遏子中由色氨酸的结合造成的构象变化。左：apo 阻遏子（PDB：3WRP）；右：当色氨酸结合到阻遏子上，识别螺旋（红色）会被移动，可以结合到两个连续的 DNA 大沟中（PDB：1TRO）。

同源框结构域蛋白形成了一个转录激活子蛋白的单体家族，这种蛋白质第一次被发现是在果蝇（*D. melanogaster*）的身体发育功能中。通过这些蛋白质识别的 DNA 序列是非常相似的，因此被称为同源框序列（homeobox sequence）。

同源框结构域是与原核生物的螺旋 - 转角 - 螺旋阻遏子结构很相似的，但不同的是它们都是单体。它们是通过三个螺旋形成的，这三个螺旋中的第二和第三个以相同的方式排列在螺旋 - 转角 - 螺旋蛋白中。与大沟的相互作用是通过第三个螺旋。一个相对较大量的非特异性相互作用使得其结合更加牢固（图 7.25）。这些非特异性的相互作用是通过 N 端的臂和前两个螺旋形成的。

为了在不同的结合位点之间形成更加有效的区分度，同源框结构域会与其他的一些转录因子共同行使功能。酵母 *MAT* 基因决定了细胞的交配型是与倍数性状态（ploidy status）相关的。在酵母阻遏子 MATα2 存在时，另外两个因子会与阻遏子作用，结合到 DNA 的另一个位点上。MAT α1 与 MAT α2 形成复合物结合到 DNA 上。MAT α1 也是一个同源框结构域蛋白，这两个蛋白质是通过疏水作用结合到一起的，接触的是 MAT α2 的第四个螺旋和 MAT α1 的一个疏水腔（图 7.25）。增加的特异性是同时结合在两个位点的结果，这需要结合位点具有正确的空间和相对的方向。

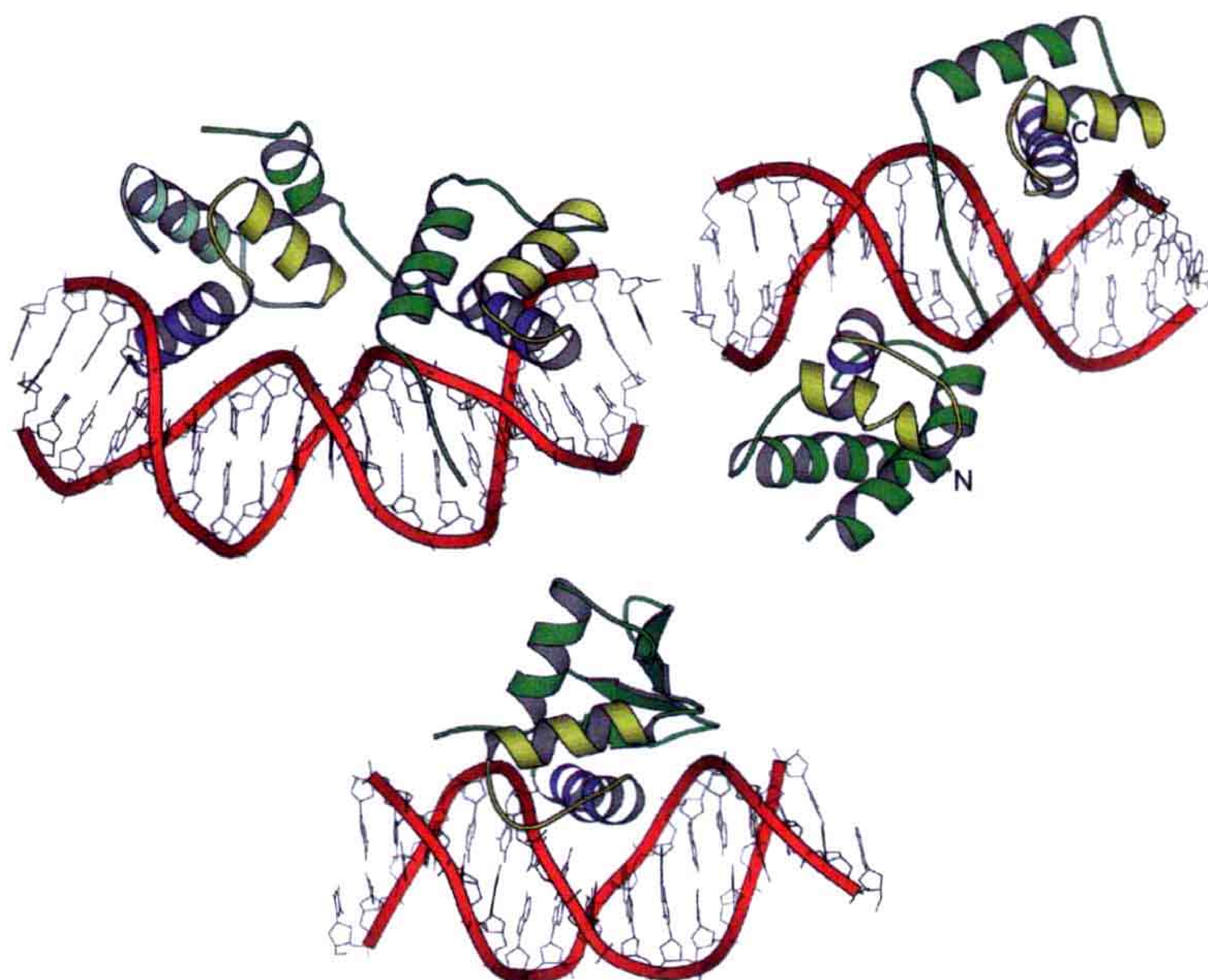


图 7.25 含有螺旋 - 转角 - 螺旋模式的真核生物转录因子。上左：MAT α1-MAT α2 复合物中同源框结构域（homeodomain）与 DNA 的结合。两个 DNA 结合结构域都含有三个螺旋，其中一个 N 端螺旋在螺旋 - 转角 - 螺旋模式之前（黄色和蓝色螺旋）。MAT α2 的一个 N 端尾巴与碱基和小沟中的结构骨架相互作用。MAT α2（绿色）还有一个螺旋是用来形成二聚体的（PDB：1LE8）。上右：人类蛋白 Oct-1 中 POU 区域的两个 DNA 结合结构域（PDB：1OCT）。同源框结构域在 DNA 的上面而 λ 类似结构域在 DNA 的下面。下：一个带侧翼的螺旋 - 转角 - 螺旋结构域，转录因子 PU.1 的 ETS 结构域结合到了 DNA 上（PDB：1PUE）。

在蛋白 Oct-1 中，DNA 结合是通过相同蛋白质中两个结构相似的结构域。在这种情况下，其中一个结构域是同源框结构域类型，而另一个要更加类似于在 λ 阻遏子中找到的螺旋 - 转角 - 螺旋构象。这两个 DNA 结合结构域结合到了 DNA 相反的两侧，连接的连接链是部分无序的（图 7.25）。

同源框结构域蛋白根据它们的氨基酸序列可以被识别，但是一般的，确定在基因组中哪里是同源框结构域结合的位点是很困难的，特异性识别最常见的是通过多于一个蛋白质实现的，一个孤立的结构对于识别它们真正的目标序列不是很有效。

带侧翼的螺旋 - 转角 - 螺旋模式在一些转录因子中被找到，它包含一个由两个 β 发夹形成的小的 β 片层，它在螺旋 - 转角 - 螺旋模式的侧面（图 7.25）。

7.6.4 锌指与 DNA 的相互作用

锌指模式在很多 DNA 结合蛋白中都能被找到。它们都是相对较小的结合了锌离子的单位（见第二章）。在经典的锌指中，两个组氨酸和两个半胱氨酸残基结合了锌离子。这个小的单位含有 30 个氨基酸残基，形成了一个独立的折叠结构域，它是由锌离子来稳定的。

这一类型的锌指经常出现在蛋白质的一前一后，一个含有经典的锌指的蛋白质的例子是哺乳动物转录因子 Zif268（较早的生长应答蛋白 1）。这个蛋白质的 DNA 结合结构域包含了三个锌指结构。Zif268 的三个结构域的结构显示了它们都是非常相似的，包含一个 β 发夹和一个 α 螺旋结构。三个模式在大沟中的排列非常相似，由来自一个螺旋末端的残基和前面的环提供的残基与碱基特异性识别（图 7.26）。

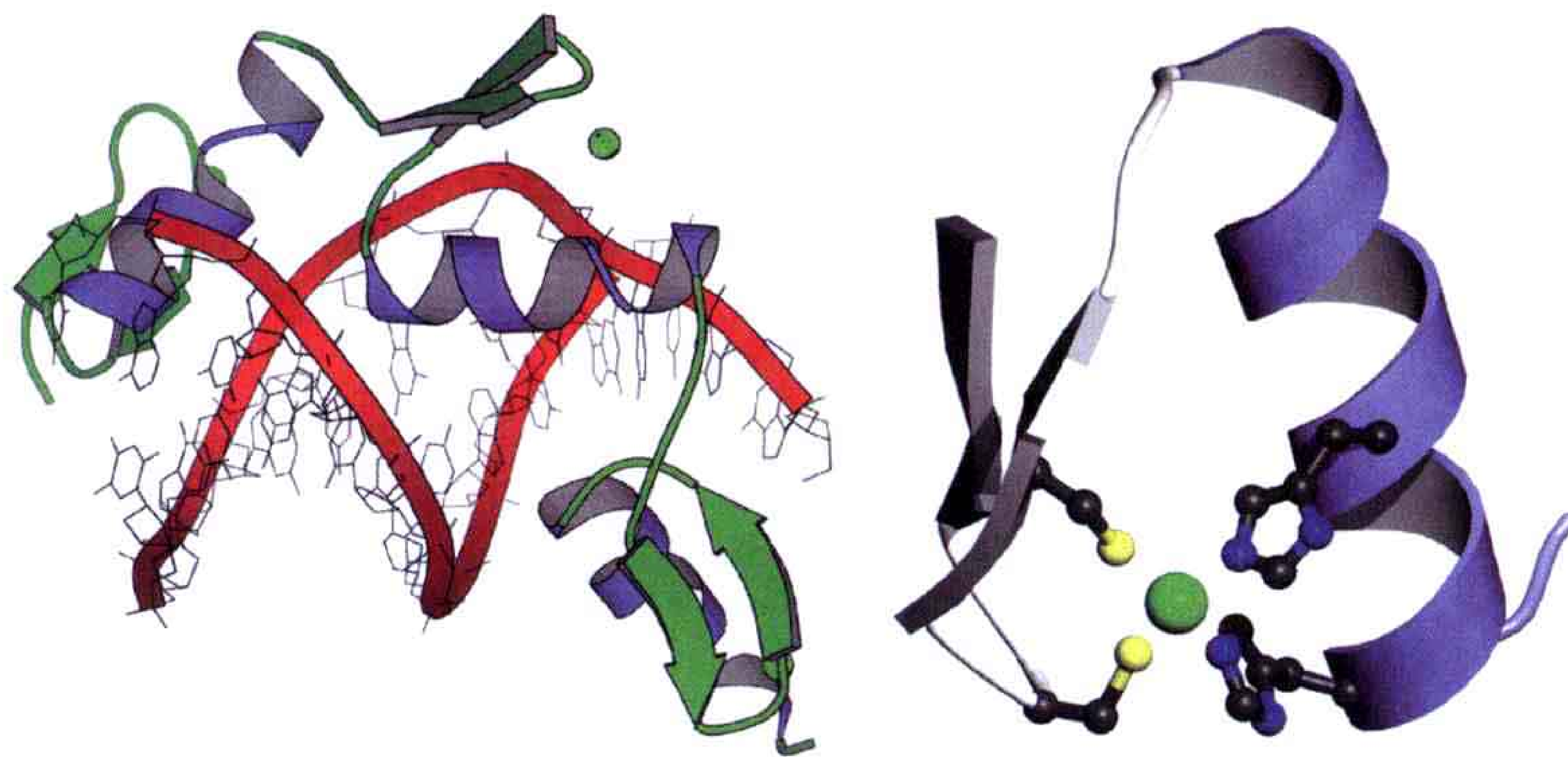


图 7.26　左：Zif268 中沿着它的 DNA 大沟的三个锌指的排列；右：细节显示了锌离子在第二个指结构中的配位情况（PDB：1ZAA）。

糖皮质激素（glucocorticoid）受体是细胞内的蛋白质，其属于一个含有保守 DNA 结合区域的蛋白家族，这个区域大约有 70 个氨基酸残基，它在葡萄糖代谢中结合皮质醇。这一家族被称为“核受体（nuclear receptor）”，它们含有另外一个可以激活受体的结构域。一些蛋白质结合甾类激素（steroid hormone），而另一些结合其他的小的疏水分子。DNA 结合结构域比经典的锌指结构要大，含有两个锌结合位点，其中每个有 4 个半胱氨酸配位到金属上。这个结构域是一个单球状的单元，包含了两个螺旋和一个环区域，并且它们通过锌离子结合在了一起。

糖皮质激素受体是一个二聚体，结合到了 DNA 的回文序列上。在复合物中，每个亚基中的一个螺旋结合到了大沟中。它们的二聚体是通过蛋白质的 C 端部分，以二聚化结合到同一侧 DNA 上的两个相邻大沟上，与二聚的原核生物阻遏子的结合很相似（图 7.27）。

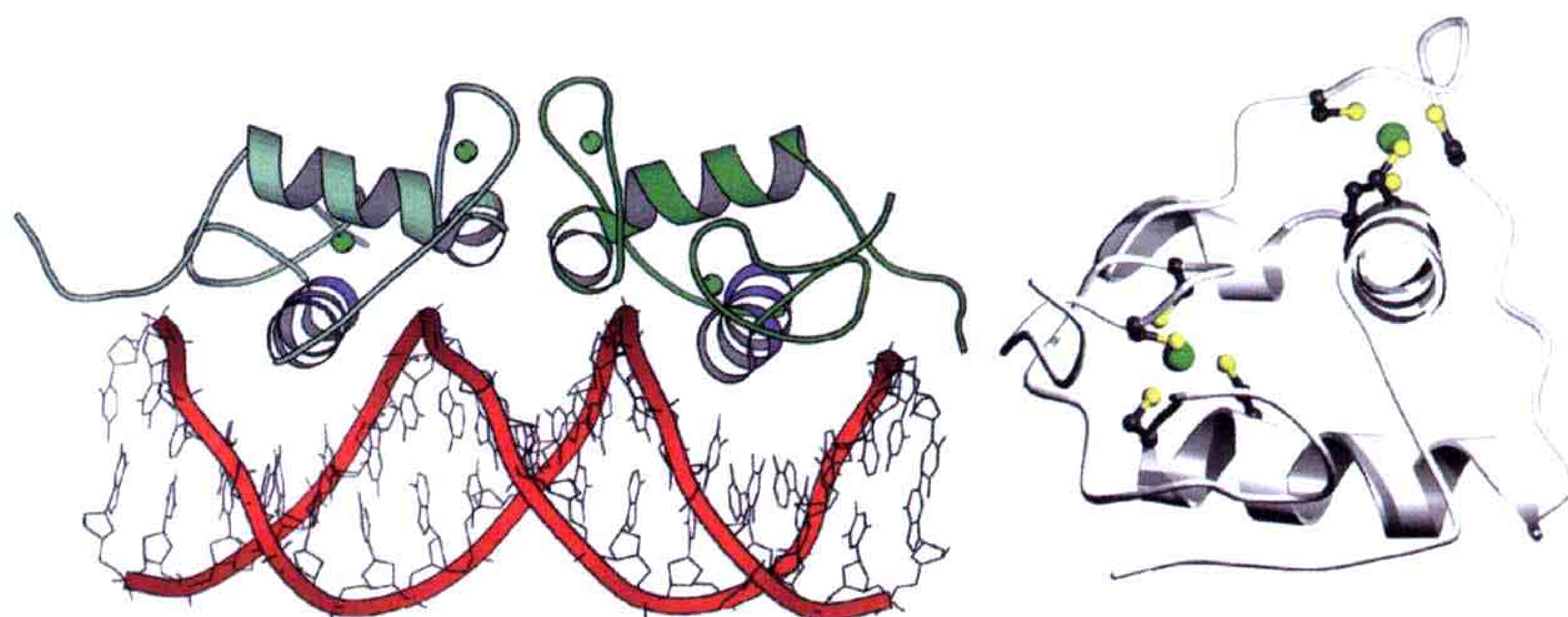


图 7.27　左：二聚的糖皮质激素受体在 DNA 上的排布；右：在受体中关于锌离子结合的细节（PDB：1GLU）。

7.6.5 p53

p53 是一个潜在的转录激活子，可以诱导很多目标基因的表达，从而间接地介导 DNA 修复，细胞增长间隙和细胞凋亡中的产物。在这种方式中，p53 能够抑制细胞的转化和肿瘤的形成。p53 突变，或者与病毒或细胞内的蛋白质联合时会失去活性，这是导致人体内差不多 50% 的癌症的原因。p53 也可以行使负调控的功能。

p53 是一个存在时间较短的蛋白质，一般情况下会保持很低的水平。在对一系列生理条件做出反应时，它会在细胞核中积累，并且变得有活性，激活一系列基因的表达。它是一个四聚体，是因为四聚化结构域的相互作用。通过 p53 的转录激活时由于它的 N 端激活结构域（TAD）具有很高的酸性，p53 的 TAD 可以与一般的转录因子 TF Ⅱ D 和 TF Ⅱ H 及其他的一些蛋白质相互作用。TAD 不同的部分具有不同的转录活性功能，会与不同的蛋白质相互作用。

DNA 结合结构域（DBD）是一个核心结构域。大约 95% 的由 p53 产生的癌症都是因为在这一结构域有了突变。p53 激活转录是通过识别启动子的共有序列：PuPuPuC(A/T)(T/A)GPyPyPy。DNA 的结合表面大部分含有三个单元（图 7.28）。一个环区和 C 端螺旋与 DNA 的大沟吻合，而另一个环提供了一个精氨酸靠近小沟。这个环和另外的一个环结合在一个锌离子上，这对它们的构象和蛋白质的结合都是非常重要的。

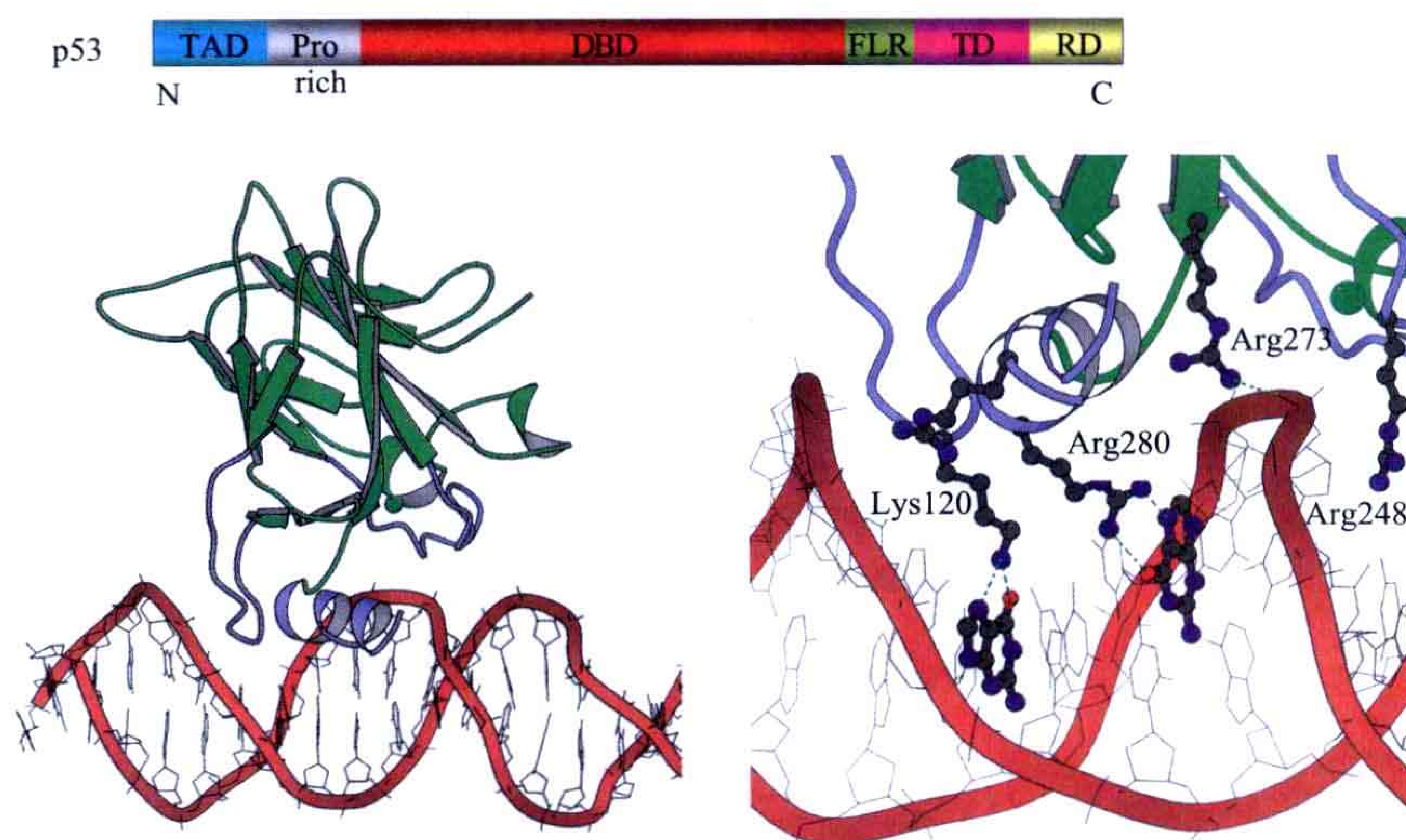


图 7.28 上：p53 的结构域搭建图。转录激活结构域（TAD）可以被广泛的磷酸化。DBD 结构域是 DNA 结合核心结构域。FLR 是一个紧接着四聚化结构域（TD）的柔性链区域。RD 是一个控制结构域。最后两个结构域可以被乙酰化。下左：p53 的 DBD 结构域与 DNA 相互作用的结构。当一个螺旋和一个环结合到大沟中时，另一个环会与毗邻的小沟相互作用。下右：与 DNA 相互作用侧链的详细表述图。Arg280 与鸟嘌呤的作用非常关键（PDB：1TUP）。Arg273 与磷酸上的氧原子相互作用，而 Arg248 像三明治一样被夹在小沟中，都是在肿瘤中经常被突变掉的。

p53 结合到一段典型的 DNA 区域，它含有 4 个相同序列的拷贝。也有可能 p53 四

聚体的 DBD 会分别结合到这 4 个接近的拷贝上。

7.6.6 结合特异性

蛋白质结合到 DNA 中的特异性序列主要依赖于侧链和碱基之间的氢键，以及在相互作用中依赖于 DNA 的骨架构象。因为 DNA 中的每个碱基对都显示了独一无二的氢键受体和供体图谱，它们可以被侧链识别，像精氨酸和谷氨酸，同样也有特定的氢键受体和供体的图谱（图 7.29）。

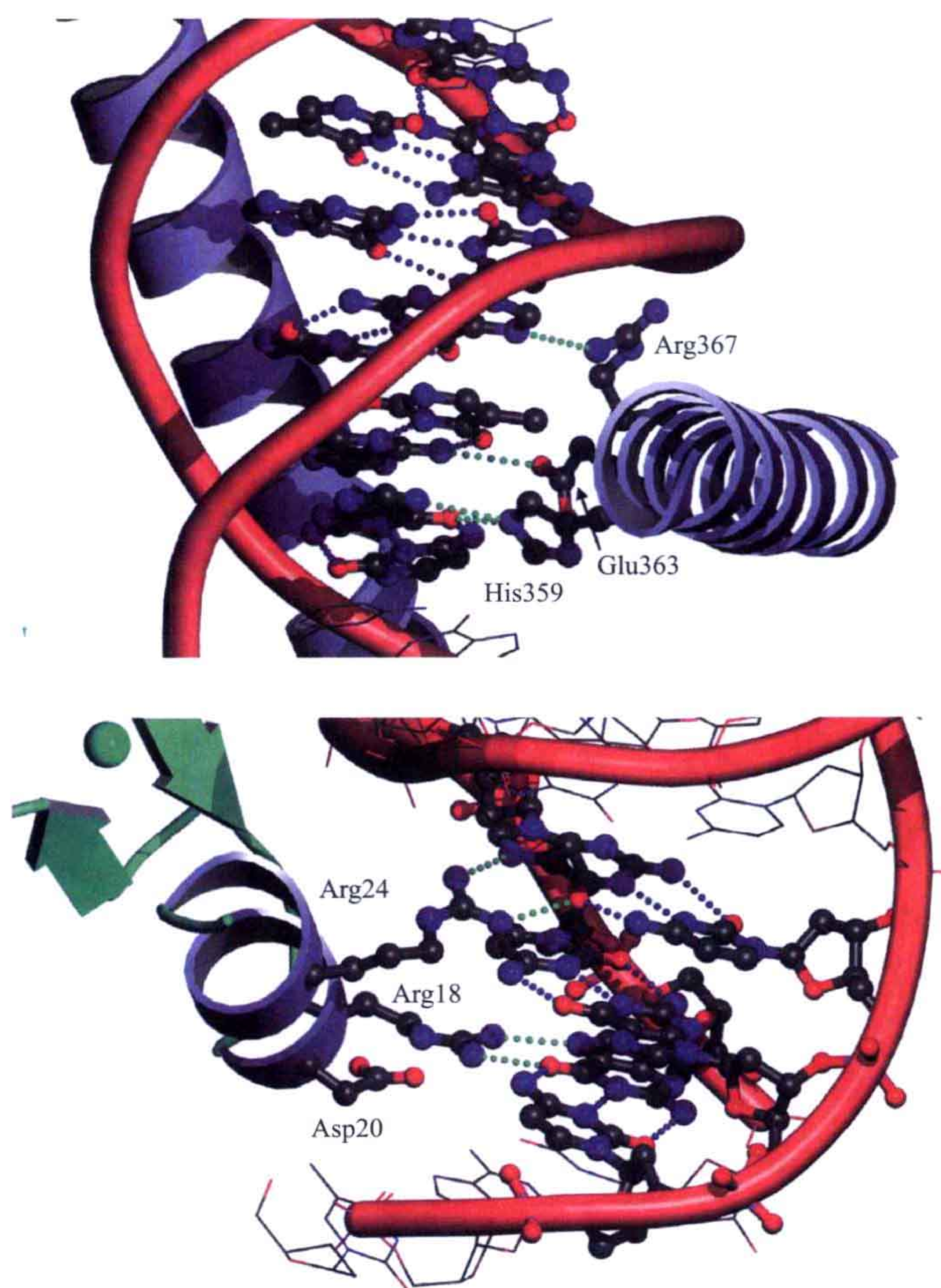


图 7.29 上：Myc 与 DNA 的相互作用（也可以见图 7.22）。异质二聚体特异性结合到一个六核苷酸片段 CACGTG 上。螺旋中的三个残基与三个碱基对特异性相互作用形成了回文序列识别的一半，而在另一个单体中相同的残基会以相似的方式与另一半序列结合(没有显示)。精氨酸是特别重要的残基，识别序列第三个位置的 CG 碱基对。蛋白质与 DNA 之间的氢键被表示成绿色；碱基对中的氢键被表示成蓝色。在 DNA 骨架中更多的相互作用稳定了这样的结合（PDB：1NKP）。下：Zif268 的第一个锌指与 DNA 的相互作用（也可以见图 7.26）。两个精氨酸残基与鸟嘌呤碱基的相互作用。Asp20 与 Arg18，使得它保持在这个位置。

尽管大部分的转录因子是利用螺旋结合到 DNA 大沟中来特异性识别的，但是也有一些蛋白质通过不同的方式识别的例子。一个例子是，来自噬菌体 P22 的 Arc 阻遏子。这个蛋白质是一个四聚体，其中一对单体形成了反平行的双链片层，结合到了 DNA 的大沟中（图 7.30）。

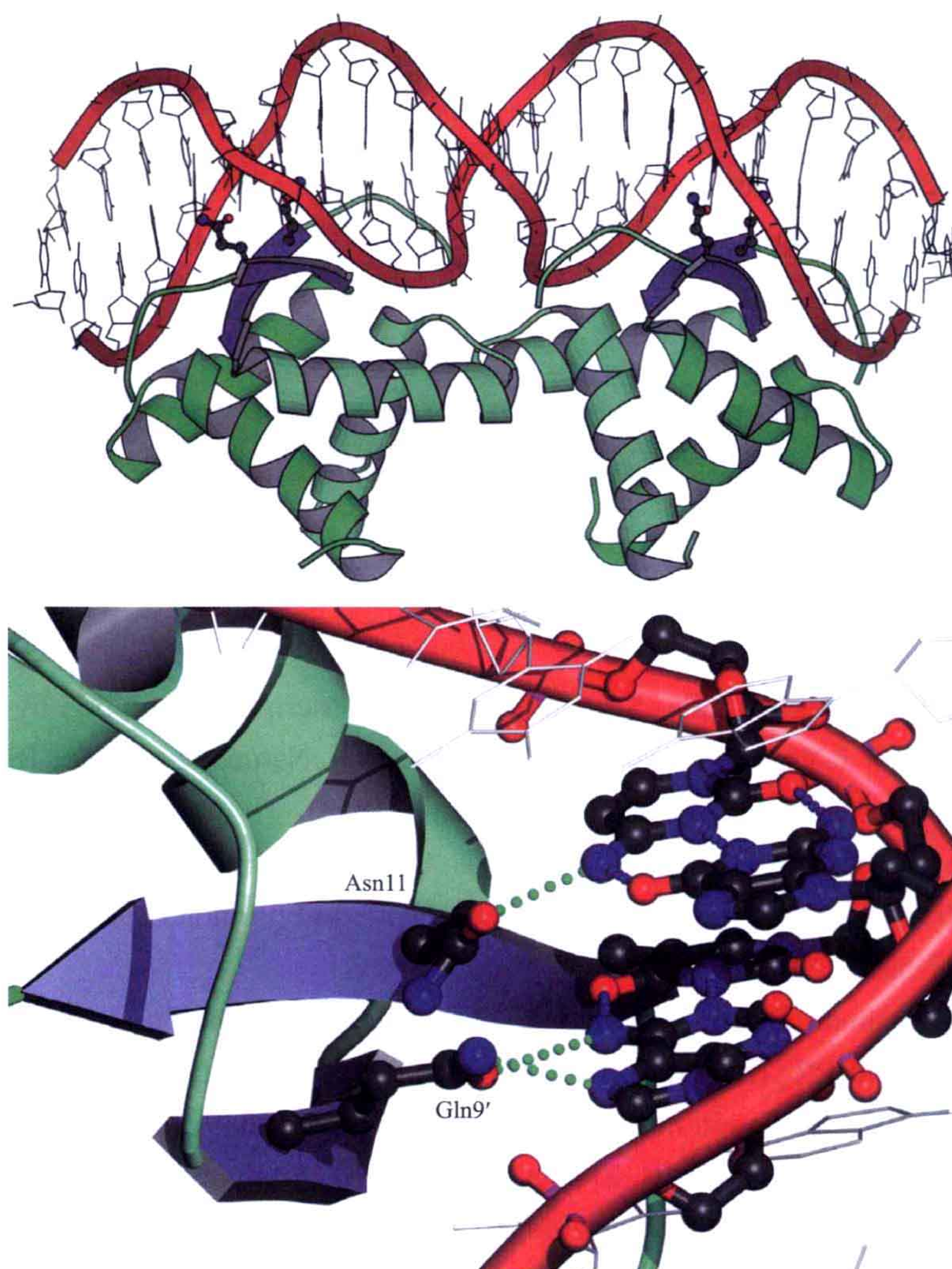


图 7.30 上：Arc 阻遏子四聚体；下：相互作用的细节显示了蛋白质是通过一个谷氨酸残基来识别腺嘌呤碱基的。一个天冬酰胺结合到了胞嘧啶的氨基基团上（PDB：1BDT）。

这些重要进展的取得需要感谢转录机器的结构鉴定，这导致了在功能方面的更加深入的理解。这是很明显的，但是，还是有很多问题需要去探索。对于一些转录因子，如 TFⅡD、TFⅡH 和媒介物都是知之甚少的。除此之外，我们还没有关于 RNA 聚合酶与 DNA 和一群转录因子组装后启动转录的有效模型。

延伸阅读（7.1 节）

综述文章

Kuo MH, Allis CD. (1998) Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays* **20:** 615-626.

Luger K. (2006) Dynamic nucleosomes. *Chromosome Res.* **14:** 5-16.

Santos-Rosa H, Caldas C. (2005) Chromatin modifiers, the histone code and cancer. *Eur J Cancer* **41:** 2381-2402.

延伸阅读（7.2 ~ 7.4 节）

原始文献

Brueckner F, Cramer P. (2008) Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase Ⅱ translocation. Nature Struct Mol Biol, Advance online publication 13 June 2008 (DOI: 10.1038/nsmb. 1458).

Campbell EA, Muzzin O, Clenov M, Sun JL, Olson A, Weinman O, Trester-Zedlitz ML, Darst SA. (2002) Structure of the bacterial RNA polymerase promoter specificity σ subunit. Mol Cell **9:** 527-539.

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maler-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg R. (2000) Architecture of RNA polymerase Ⅱ and implications for the transcription mechanism. Science **288:** 640-649.

Vassylyev DM, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. (2007) Structural basis substrate loading in bacterial RNA polymerase. Nature **448:** 163-169.

Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg R. (2006) Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. Cell **127:** 941-954.

综述文章

Cramer P. (2002) Multisubunit RNA polymerases. Curr Op Struct Biol **12:** 89-97.

Cramer P. (2004) RNA polymerase Ⅱ structure: From core to functional complexes. Curr Op Gen Develop **14:** 218-226.

Carmer P, Armache K-J, Baumli S, Benkert S, Brueckmer F, Buchen C, Damsma GE,

Deng IS, Geiger SR, Jasiak AJ, Jawhari A, Jenneback S, Kamenski T, Kettenberger H, Kuhn C-D, Lehmann E, Leike K, Sydow JF, Vannini A. (2008). Structure of eukaryotic RNA polymerases. Annu Rev Biophys **37:** 337-352.

Kornberg RD. (1999) Eukaryotic transcription. Trends Biochem Sci **24:** M46-M49.

延伸阅读（7.5 节）

原始文献

Bushnell DA, Westover KD, Davis RE, Kornberg RD. (2004) Structural basis of transcription: An RNA polymerase Ⅱ -TF Ⅱ B cocrystal at 4.5 Å. Science **303:** 983-988.

Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG, Burley SK. (1995). Crystal structure of a TF Ⅱ B-TBP-TATA-element ternary structure. Nature **377:** 119-128.

综述文章

Chadick JZ, Asturias FJ. (2005) Structure of eukaryotic Mediator complexes. Trends Biochem Sci **30:** 264-271.

Deng W, Roberts SGE. (2007) TF Ⅱ B and the regulation of transcription by RNA polymerase Ⅱ . Chromosoma **116:** 417-429.

延伸阅读（7.6 节）

原始文献

Canadillas JMP, Tidow H, Freund SMV, Rutherford TJ, Ang HC, Fersht AR. (2006) Solution structure of p53 core domain: Structural basis for its instability. Proc Natl Acad Sci USA 103: 2109-2114.

Cho Y, Gorina S, Jeffrey PD, Pavletich NP. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. Science 265: 346-355.

（金坚石 译 曹 駸 苏晓东 校）

第八章

蛋白质的合成 —— 翻译

8.1 翻译系统的演化

从遗传信息到功能蛋白分子的翻译是生命活动的中心过程。遗传密码、转运 RNA（tRNA）分子及蛋白质合成机构都是高度保守的。核糖体是翻译过程发生的场所，由蛋白质和 RNA 分子组成。

卡尔·伍斯（Carl Woese）发现来自大量不同物种间的核糖体 RNA（rRNA）分子片段序列是彼此相关的。因此，核糖体 RNA 序列可以用来分析物种间的演化关系。随后，他又提出当时将生物界分为原核生物（prokaryote）和真核生物（eukaryote）两类的分类方法是不正确的，而应该将古细菌（archaea）作为一个单独的全新种类引入分类体系中。根据伍斯的观点，生物界应该被分为细菌、古细菌和真核生物三类。

对已测序的全基因组进行比较，发现翻译机构所需的分子组分占据了生命体中保守组分的大多数。遗传密码、tRNA、rRNA、核糖体蛋白和翻译因子一定是在生命演化过程的早期共同演化而来的，并在随后的演化过程中经历了有限的变化。因此，参与翻译过程的分子和翻译机制可以为我们提供这样一个信息，即现代细胞的祖先是如何进行更为原始的蛋白质合成的。

在与大多由蛋白质占据主导地位的细胞活动进行比较时，蛋白质合成过程的一个重要特点是核酸分子扮演了中心角色，这些角色包括信使 RNA (mRNA)、tRNA 和核糖体 RNA 分子，mRNA 包含基因序列的拷贝并结合到核糖体上，tRNA 分子负责解码基因序列并将相应的氨基酸连接到正在核糖体中合成的多肽链上。

8.1.1 遗传密码和 tRNA

组成信使 RNA（mRNA）遗传密码（图 8.1）的碱基三联体称为密码子（codon）。密码子对应 20 种不同的氨基酸。此外，通常还有三种终止密码子（stop codon），即 UAA、UAG 和 UGA。起始密码子与编码甲硫氨酸的密码子（AUG）相同。生物体有特殊系统用于区分普通的甲硫氨酸密码子和起始密码子。有的氨基酸只有一个密码子编码，如甲硫氨酸和色氨酸，而有的氨基酸却有多达 6 个密码子编码，如丝氨酸、亮氨酸和精氨酸。这反映了密码子的简并性（degeneracy），即密码子与 tRNA 分子间并

非一一对应的关系。哺乳动物的线粒体只有非常有限的 tRNA 分子种类，但很多物种有大约 40 种。这关系到密码子的使用偏好性及一些 tRNA 可以读取几个密码子的可能性（即 tRNA 的摆动配对，tRNA wobble base-paring）。

密码子第二位碱基

密码子第一位碱基（左列）；密码子第三位碱基（右列）

	U	C	A	G	
U	Phe F	Ser S	Tyr Y	Cys C	U
	Phe F	Ser S	Tyr Y	Cys C	C
	Leu L	Ser S	STOP	STOP	A
	Leu L	Ser S	STOP	Trp W	G
C	Leu L	Pro P	His H	Arg R	U
	Leu L	Pro P	His H	Arg R	C
	Leu L	Pro P	Gln Q	Arg R	A
	Leu L	Pro P	Gln Q	Arg R	G
A	Ile I	Thr T	Asn N	Ser S	U
	Ile I	Thr T	Asn N	Ser S	C
	Ile I	Thr T	Lys K	Arg R	A
	Met M	Thr T	Lys K	Arg R	G
G	Val V	Ala A	Asp D	Gly G	U
	Val V	Ala A	Asp D	Gly G	C
	Val V	Ala A	Glu E	Gly G	A
	Val V	Ala A	Glu E	Gly G	G

图 8.1　普适性的遗传密码表。三核苷酸密码子对应的 20 种氨基酸同时用三字母和单字母简写表示。

转运 RNA 分子通常由大约 75 个核苷酸组成（见 3.3.3.1 节），呈现由一条“茎（stem）”和三片“叶（leaf）”组成的三叶草形的二级结构（图 3.52）。“茎”的 3′ 端含有一个 CCA 的序列，正是该序列中 A 的核糖被特异性的 tRNA 合成酶识别并氨酰化，这条“茎”也因而被称为氨酰茎或接受茎。三叶草的三片叶子分别被称为 D 环 (D stem and loop)、反密码子环（anticodon stem and loop）及 T 环（T stem and loop）。此外还有一个可达 21 个核苷酸长度的可变环（variable loop, V loop）。丝氨酸和亮氨酸及细菌和叶绿体中酪氨酸的 tRNA 通常有较长的可变环。位于反密码子环中部的反密码子与 mRNA 中的密码子配对。整个 tRNA 分子的三维立体结构呈现独特的“L”形结构（图 8.2）。接受茎和 T 环组成“L”中较长的一笔，而反密码子环和 D 环组成“L”中较短的一笔。T 环和 D 环相互作用形成“L”的拐点。令人吃惊的是，tRNA 分子的两个功能位点分别位于其三维结构的相反的两端，其相互距离大约为 75Å。

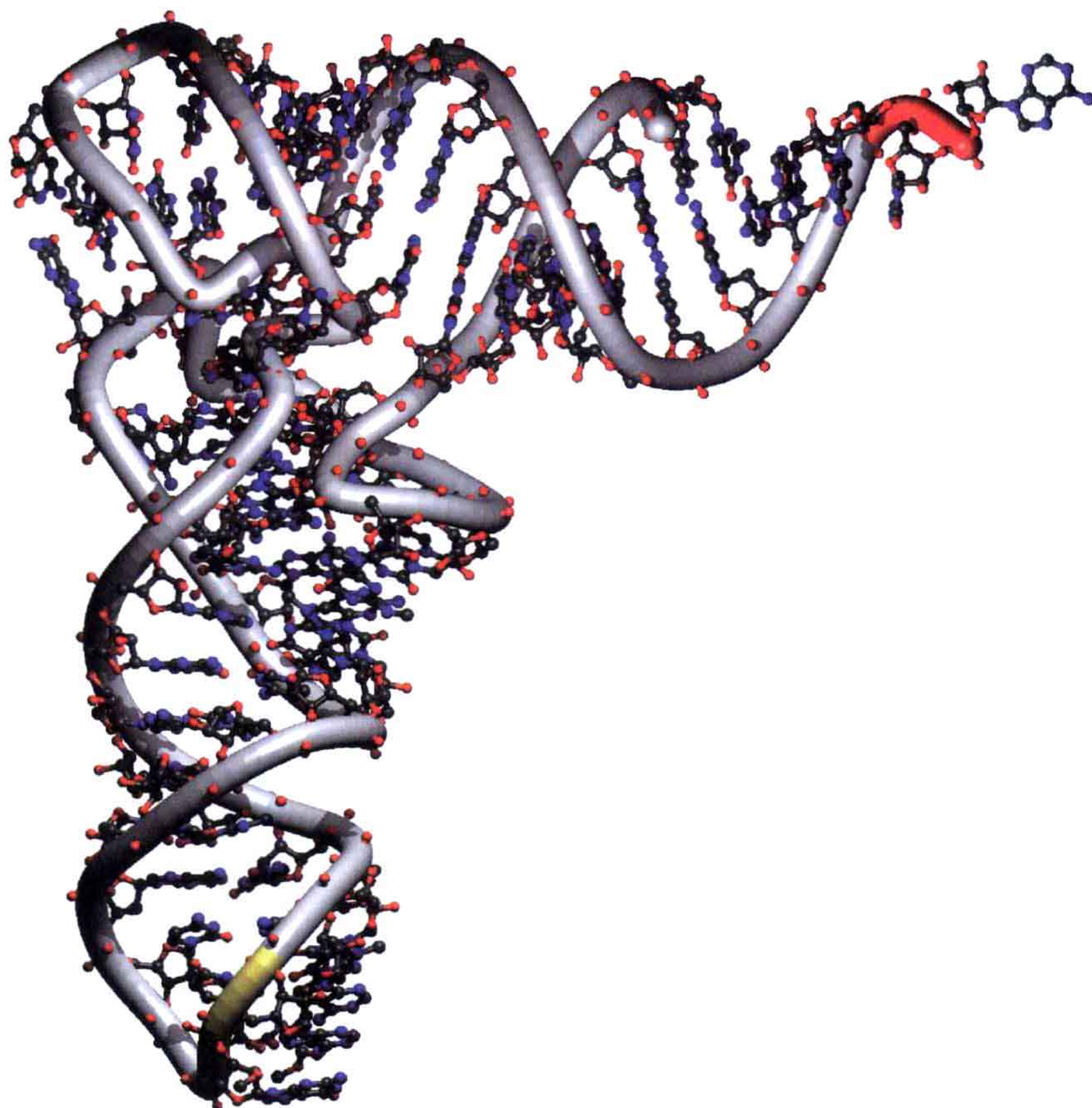
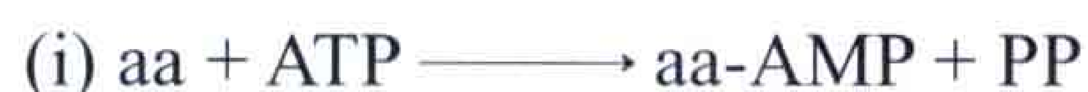


图 8.2　一个 tRNA 分子的构象。接受茎（红色区域为 CCA）和反密码子臂（黄色区域为反密码子）分别位于 L 形结构的两端；T 臂和 D 臂构成 L 形结构的拐角（PDB：1EHZ）。

8.2　tRNA 合成酶

负责氨基酸 (aa) 与 tRNA 分子结合的氨酰 tRNA 合成酶（aminoacyl-tRNA synthetase）（aaRS 或者 RS）具有氨基酸特异性，20 种氨基酸通常需要 20 种不同的 tRNA 合成酶。tRNA 合成酶分子参与的反应分两步进行：



(i) aa + ATP $\longrightarrow$ aa-AMP + PP



(ii) aa-AMP + tRNA $\longrightarrow$ aa-tRNA + AMP

氨基酸首先必须由正确的 tRNA 合成酶识别，结合后由一分子 ATP 激活形成中间体 aa-AMP，随后，在反应 (ii) 中，氨基酸被连接到同样结合该 tRNA 合成酶的 tRNA 分子上。tRNA 中的保守序列 CCA 中的末端腺苷的 2′ 或 3′ -OH 基团直接进攻 aa-AMP 的高能酯键，使氨基酸与核糖相连。翻译过程的忠实度主要取决于 tRNA 合成酶的特异性，氨基酸与 tRNA 一旦发生错配将不会在随后的过程中被修复，因此，tRNA

合成酶对其特异氨基酸的识别必须具有极高的准确性。对于一些在大小和结构上非常相似的氨基酸，tRNA 合成酶往往具备特殊的机制对它们加以区分。

8.2.1 tRNA 合成酶的分类

氨酰 tRNA 合成酶的分子质量和寡聚状态有着极大的不同（表 8.1）。根据三维结构和序列相似性可以把 tRNA 合成酶分为两大类，每一类中各有 10 种 tRNA 合成酶。其中Ⅰ型 tRNA 合成酶通常都是单体，而Ⅱ型 tRNA 合成酶总是二聚体或四聚体。

表 8.1 氨酰 tRNA 合成酶的寡聚状态

Ⅰ型											
RS	L	I	V	C	M	R	E	Q	K	Y	W
寡聚状态	α	α	α	α	α_2	α	α	α	α	α_2	α_2
Ⅱ型											
RS	S	T	G	A	P	H	D	N	K	F	
寡聚状态	α_2	α_2	$(\alpha\beta)_2$	α_4	α_2	α_2	α_2	α_2	α_2	$(\alpha\beta)_2$	

两类 tRNA 合成酶具有完全不同的结构，Ⅰ型 tRNA 合成酶的 ATP 结合结构域带有平行 β 束（parallel β-strand）的 Rossmann 折叠（图 8.3 和图 8.4），而Ⅱ型中的相应结构域则由反平行 β 束（anti-parallel β -strand）组成（图 8.5）。两类 tRNA 合成酶具有不同的共有序列（表 8.2）。两类 tRNA 合成酶识别 tRNA 上相对的位点并分别把氨基酸连接在 tRNA 末端核糖的 2′-OH（Ⅰ型）和 3′-OH（Ⅱ型）上。

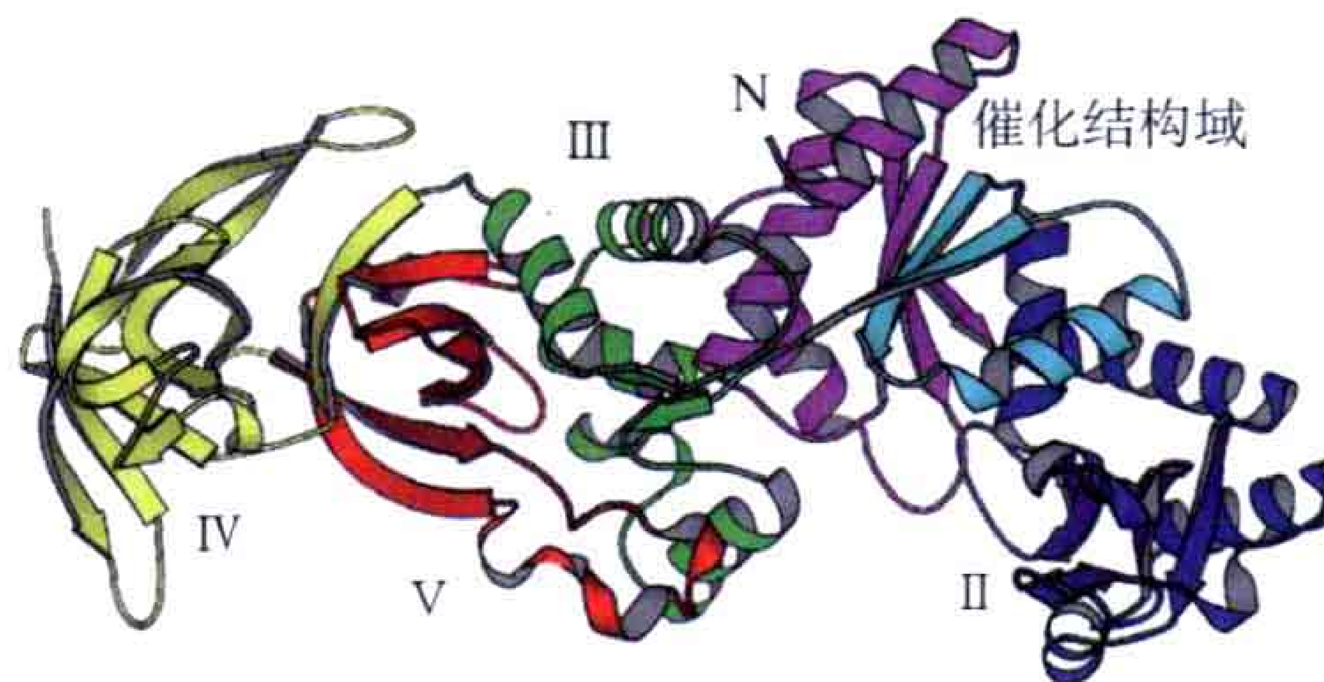


图 8.3 谷氨酰胺 tRNA 合成酶的结构，其属于Ⅰb 型。结构域Ⅱ（暗蓝色）插入催化结构域（紫色和亮蓝色）；催化结构域带有特征性的 Rossmann 折叠，活性位点位于 C 端的 β 片层；结构域Ⅲ由 α 螺旋组成，结构域Ⅳ和Ⅴ由反平行 β 桶组成；结构域Ⅲ和Ⅳ由结构域Ⅴ的一个 β 束相连（PDB：1GSG）。

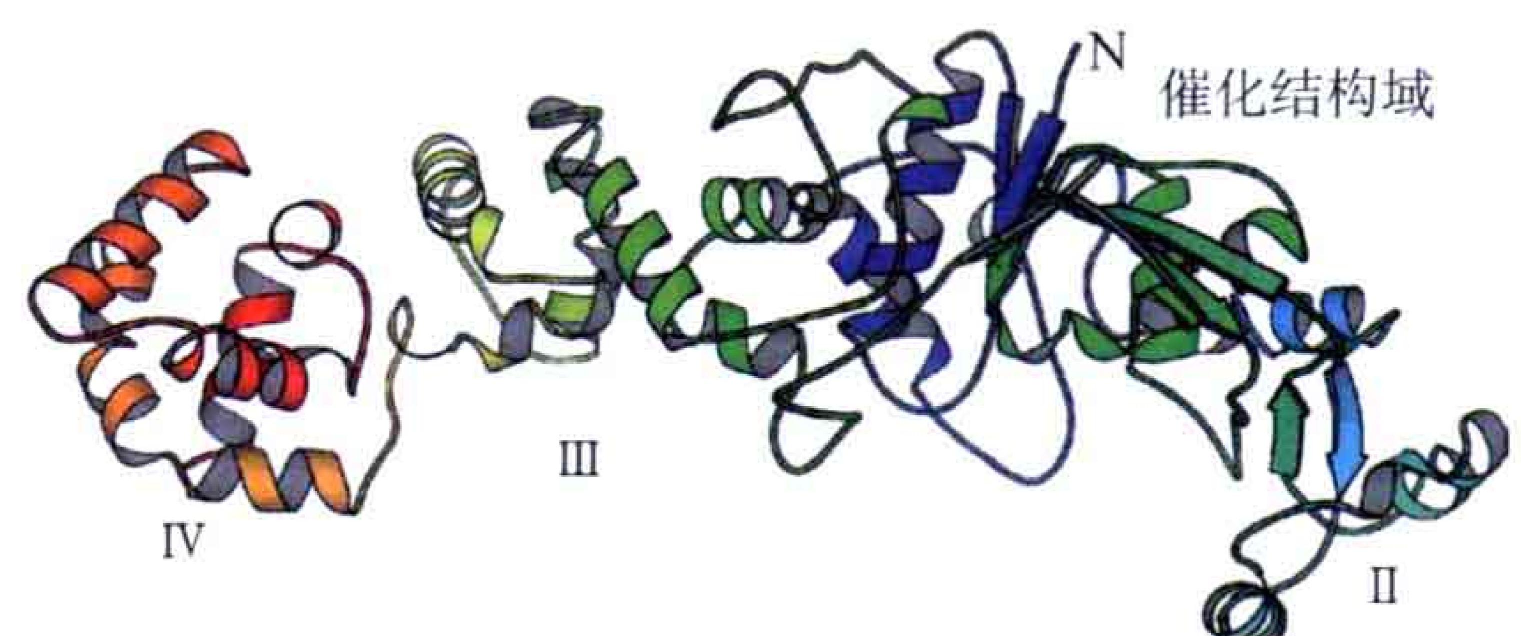


图 8.4 谷氨酸 tRNA 合成酶的结构。虽然该酶与谷氨酰胺 tRNA 合成酶属于相同的类型和亚型，但二者仍具有显著不同的结构特征。例如，谷氨酸 tRNA 合成酶的两个 C 端结构域均由螺旋结构组成（PDB：1GLN）。

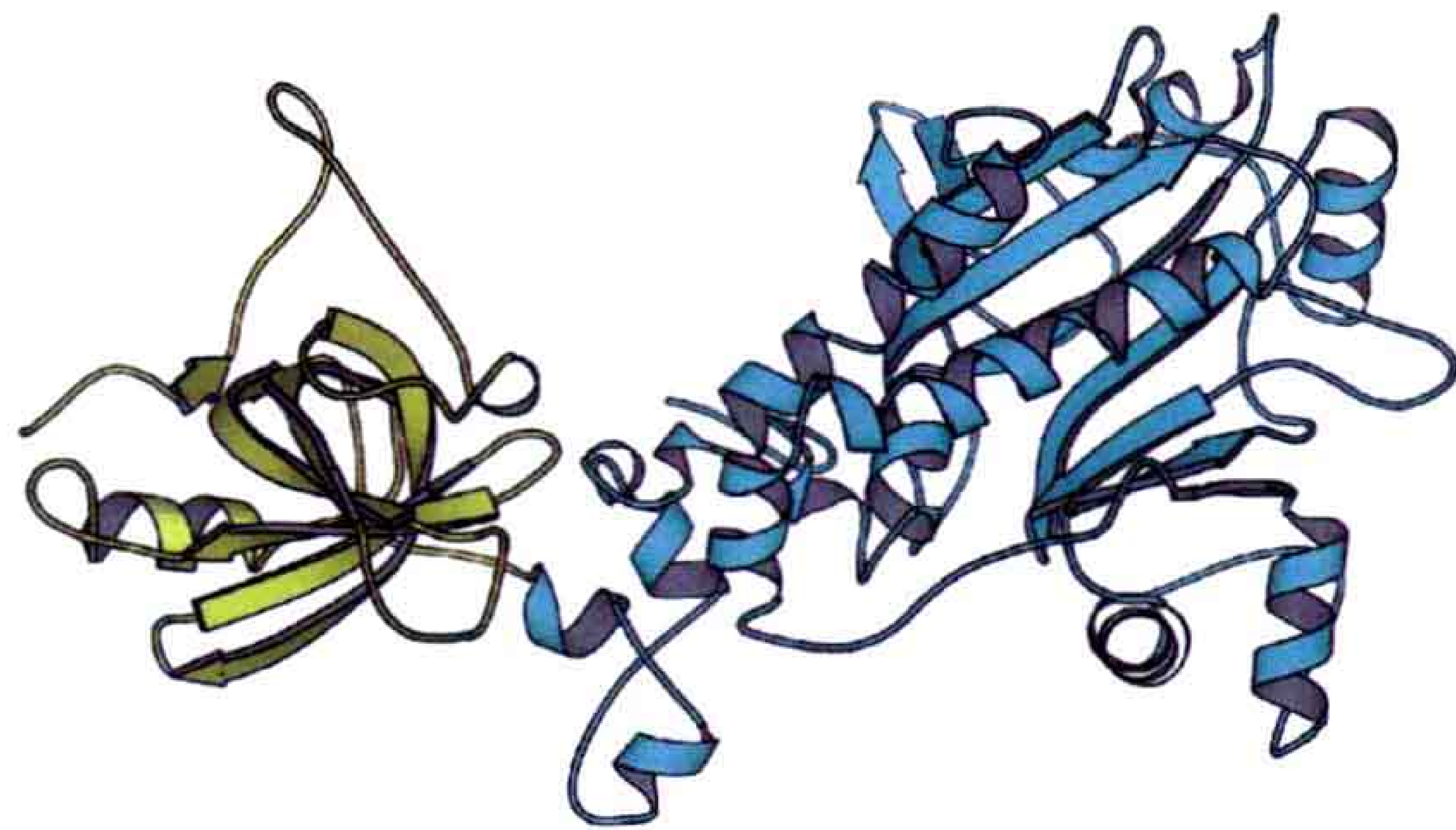


图 8.5 天冬氨酸 tRNA 合成酶的结构。该酶属于 II 型，它的催化结构域（蓝色）的特点是具有一个反平行 β 片层（PDB：1ASY）。

除了 ATP 结合结构域，两类 tRNA 合成酶在其他结构域上也表现出不同的特点。因此，根据序列同源性和结构域样式的不同，两类 tRNA 合成酶还可以进一步分成 a、b、c 三个亚型。tRNA 合成酶由模块化的结构域排布组成，而正是其中的 ATP 结合结构域的不同定义了不同的 tRNA 合成酶类型，每一个亚型中的 tRNA 合成酶都有相似的结构域排布。但是，正如图 8.3 和图 8.4 所示的，即使处于同一亚型的酶也有明显的差别。例如，谷氨酰胺 tRNA 合成酶（GlnRS）和谷氨酸 tRNA 合成酶（GluRS），虽然都有相关的结构域插入催化结构域，在催化结构域之后的结构也极为相似，但是两个酶各自的 C 端的结构域却有着显著差别——在 GlnRS 中都是由 β 桶结构组成，而在 GluRS 中都是由螺旋结构组成。

表 8.2 tRNA 合成酶的特点

		I 型	II 型
序列		HIGH	FRXE/D
		KMSKS	R/HXXXF
		GXGXGXER	
亚型	a	L, I, V, C, M, R	S, T, G, A, P, H
	b	E, Q, K	D, N, K
	c	Y, W	F
氨酰化位点		2′-OH	3′-OH
ATP 结构域折叠方式		Rossmann (平行 β)	反平行 β 折叠
氨基酸结合位点		表面	口袋深处
tRNA 接受末端		弯曲	直

8.2.2 与 tRNA 的结合

转运 RNA 上具备大量有特征的识别因子（recognition element），可以帮助 tRNA 合成酶识别相应的（cognate）tRNA（图 8.6）。两类 tRNA 合成酶结合在 tRNA 上相对的两个位点上，大多数识别因子定位在 tRNA 面对 tRNA 合成酶的一侧。在大多数情况下 tRNA 合成酶通过接受臂和反密码子识别不同的 tRNA（图 8.7），识别反密码子的中间碱基（35 位）是非常普遍的。

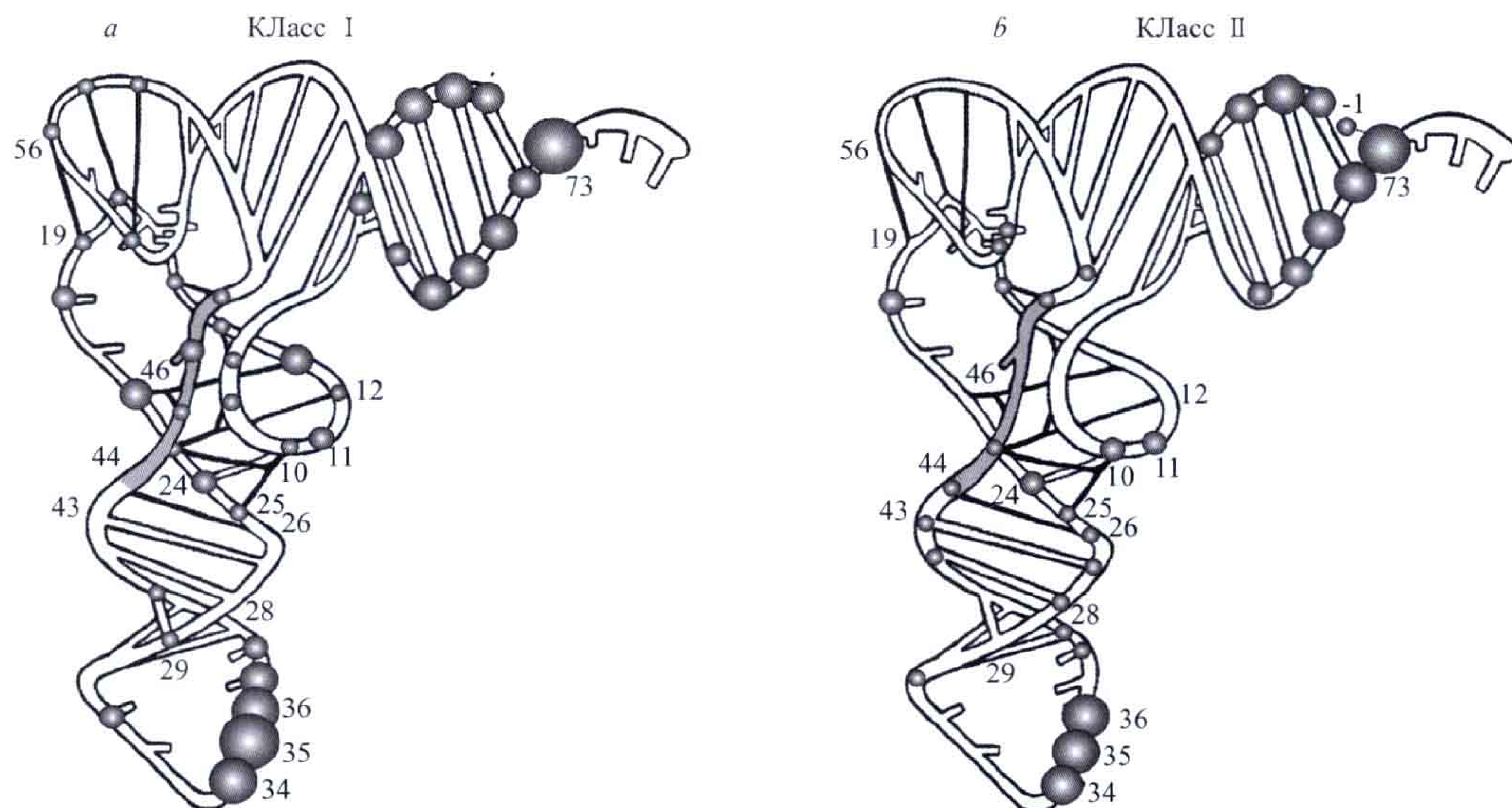


图 8.6　I 型（左）和 II 型（右）tRNA 合成酶对 tRNA 上识别元素的分布。在丝氨酸和酪氨酸 tRNA 合成酶中识别的可变臂（灰色部分）可长达 21 个核苷酸 [经 Vasileva IA, Moor NA 授权再印制 (2007) Interaction of aminoacyl-tRNA synthetases with tRNA: Genaral principles and distinguishing characteristics of the high-molecular-weight substrate recognition. *Biochemistry* (*Moscow*) **72:** 247-263.　版权 (2007) Pleiades Publishing, Inc.]

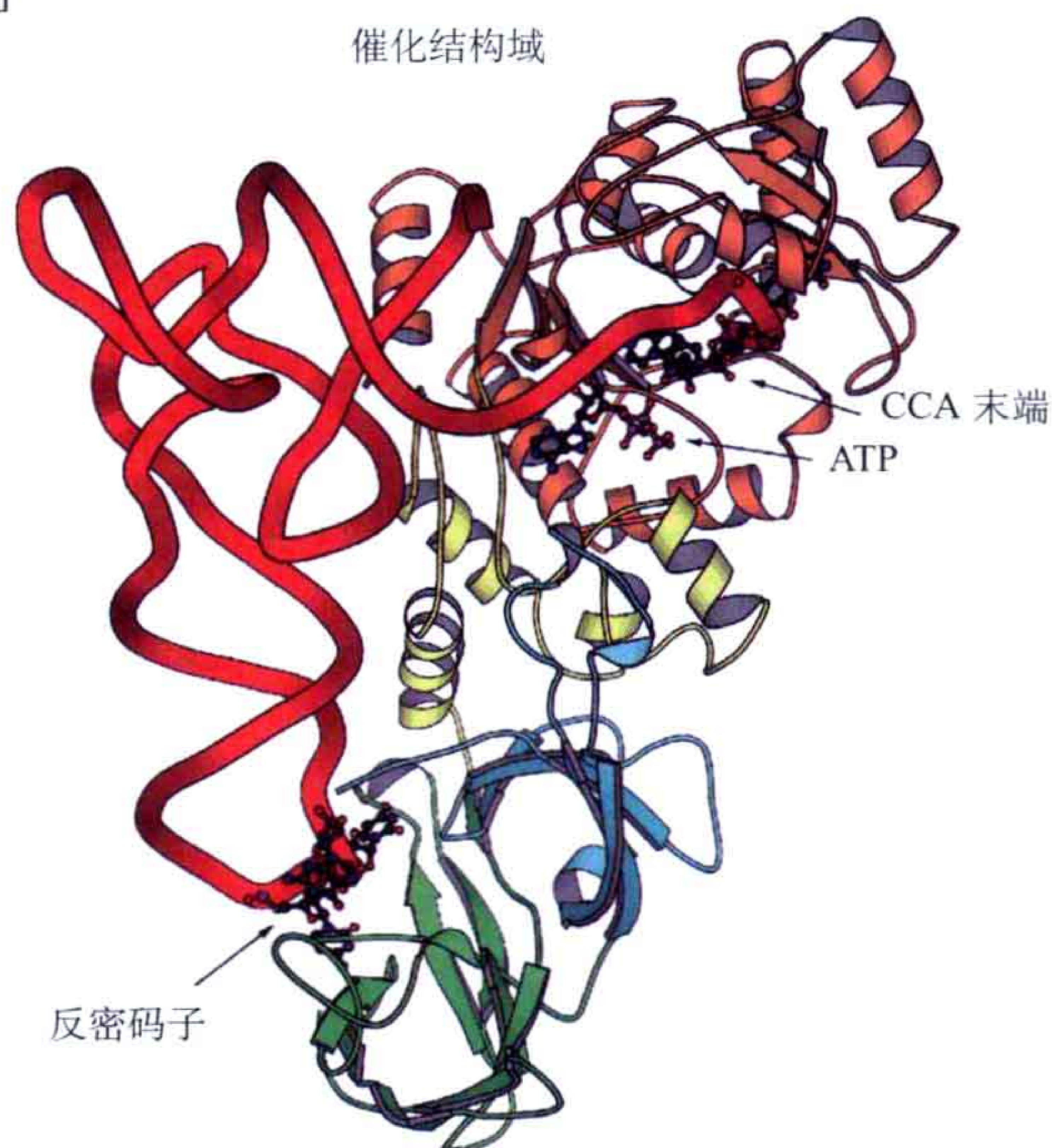


图 8.7　谷氨酰胺 tRNA 合成酶与 tRNA 的复合体结构。接受臂末端和反密码子均和酶有相互作用（PDB: 1GTR），很明显地可以看到 ATP 与弯曲的 CCA 末端非常接近。

因为氨基酸被连接在 tRNA 上相对于反密码子的一边，tRNA 合成酶需要同时与 tRNA 的两端相互作用。在天冬氨酸 tRNA 合成酶中，反密码子与酶的反密码子结合结构域相互作用，并由它们之间的氢键决定结合的特异性，然后酶通过与整个 RNA 骨架的堆积和接触进一步稳定了酶与 RNA 之间的相互作用（图 8.8）。

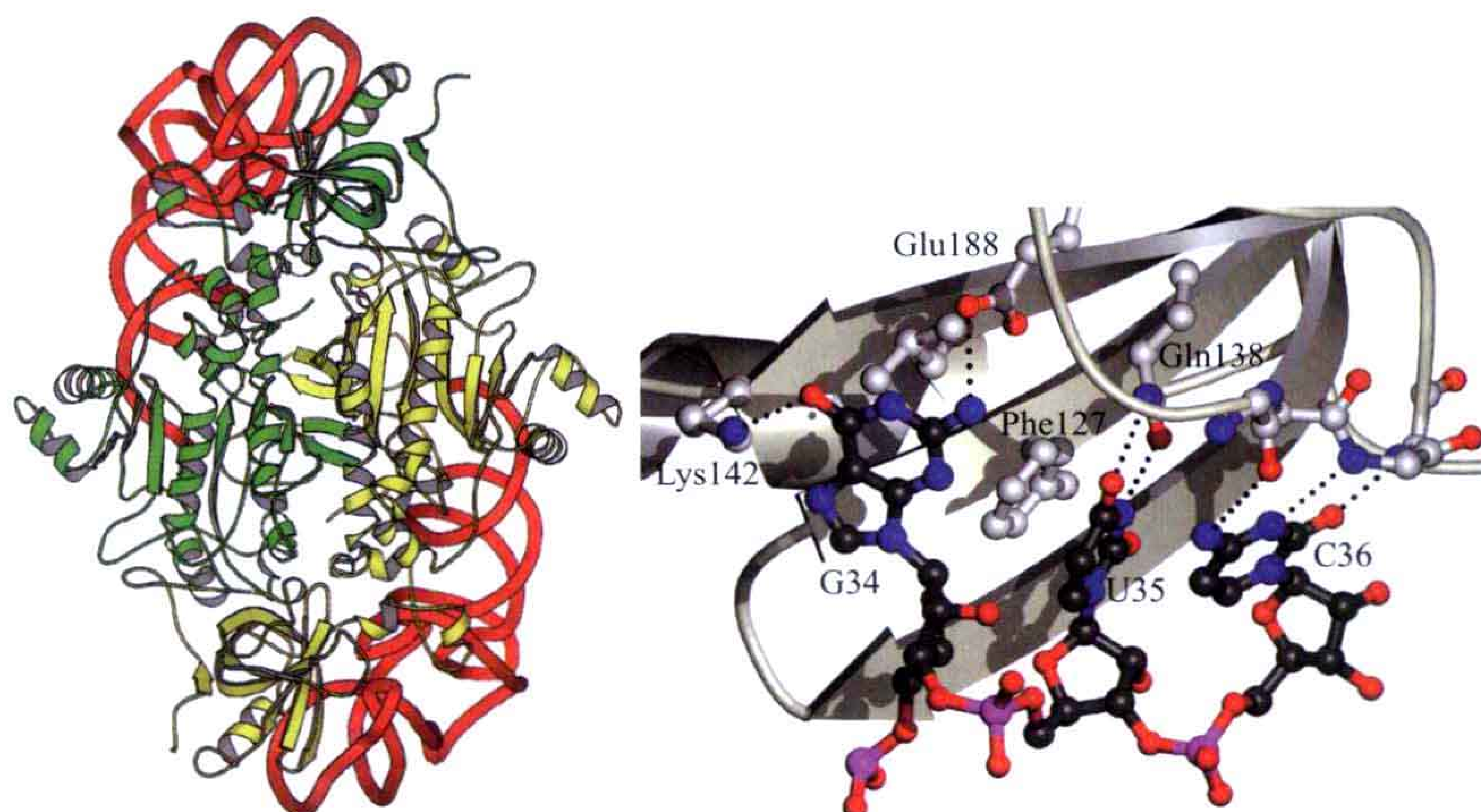


图 8.8　左：结合了两分子 tRNA 的天冬氨酸 tRNA 合成酶二聚体（Ⅱ型）；右：天冬氨酸 tRNA 的反密码子 GUC 与天冬氨酸 tRNA 合成酶的保守残基间的相互作用。从右图可以看到氢键（Lys142、Glu188、Gln138、主链）与堆积作用（Phe127），氢键用虚线表示。所有的Ⅱ b 型的 tRNA 合成酶都是由 Phe127 和 Gln138 识别反密码子中间一位的尿嘧啶（PDB：1ASY）。

在丝氨酸、丙氨酸、亮氨酸和甘氨酸 tRNA 合成酶中，反密码子并不参与酶与 tRNA 间的相互作用。丝氨酸和亮氨酸都各有 6 种不同的密码子，亮氨酸密码子的第二位碱基总是 U，但丝氨酸所有的碱基都可以有所不同，因此酶很难通过与反密码子的相互作用对其加以区分。然而，亮氨酰和丝氨酰 tRNA 都有一个共同的特点，即它们都有一条非常长的可变臂（variable arm）。丝氨酸 tRNA 合成酶属于Ⅱ型，它与 tRNA 的复合体的结构显示该酶具备一条延长的结合在 tRNA 的 TψC 臂和可变臂之间的螺旋发夹结构，该部分结构在没有 tRNA 存在的情况下是无序的。该酶的反密码子臂指向远离酶的方向，而接受臂紧密结合在酶二聚体的另一个亚基的活性位点上。对于亮氨酸 tRNA 合成酶，不同的 $tRNA^{Leu}$ 的较长的 V 环（V loop）可以被古细菌中的酶（archaeal enzyme）识别（图 8.9），但是不能直接被细菌中的酶（bacterial enzyme）识别。

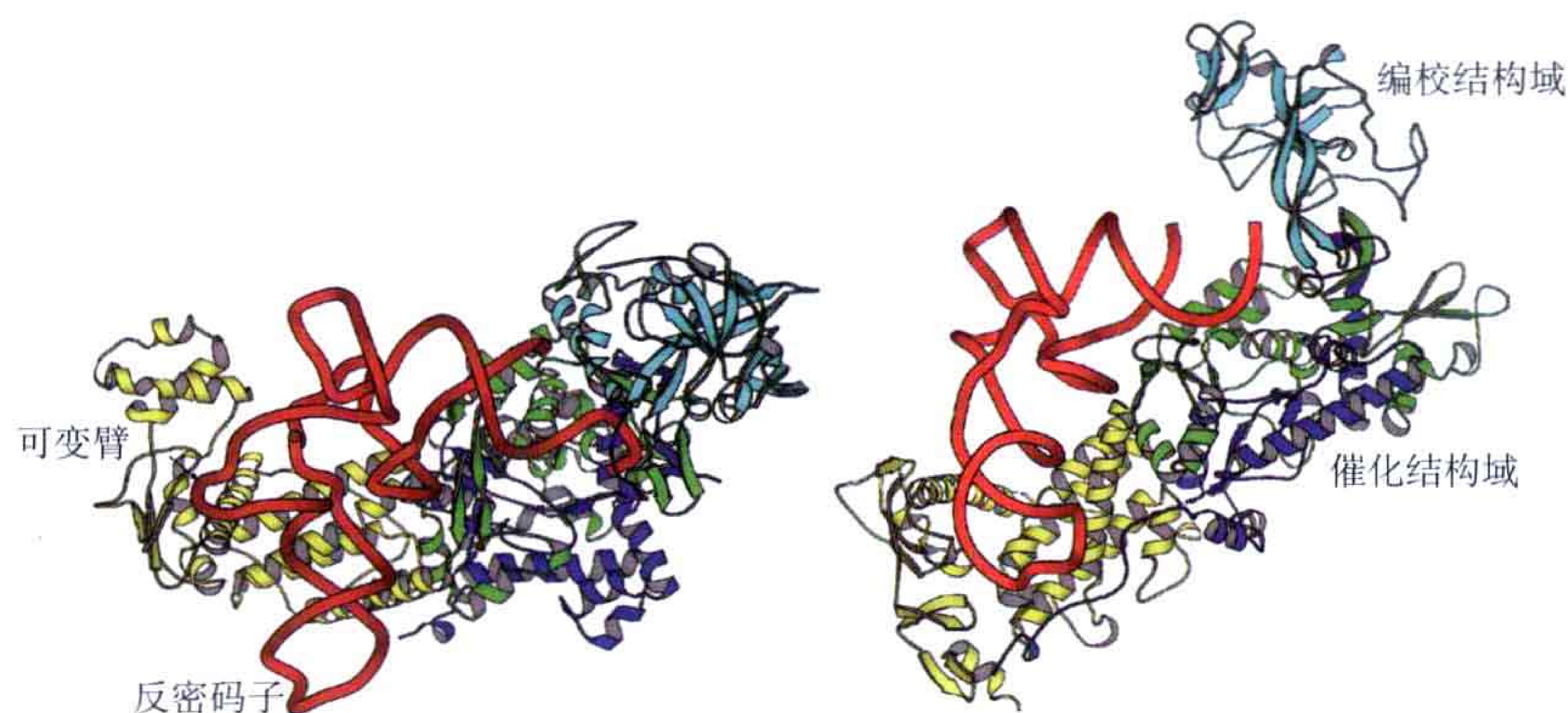


图 8.9　左：亮氨酸 tRNA 合成酶，该酶是一个不直接结合在反密码子上的例子，而是部分依赖于 $tRNA^{Leu}$ 具有的比其他 tRNA 更长的可变臂（PDB：1WZ2）。右：异亮氨酸 tRNA 合成酶的结构显示了编校结构域的作用，当结合到催化或编校结构域时，接受臂会发生变化。注意该例中反密码子参与了对 tRNA 的识别（PDB：1QU2）。

8.2.3　对氨基酸的选择——编校

不仅仅是 tRNA 会难以区分，一些氨基酸在结构和本质上也是非常相似的。例如，缬氨酸可以被很好地放到异亮氨酸 tRNA 合成酶的口袋中；苏氨酸可以放到缬氨酸 tRNA 合成酶的口袋中；而丝氨酸可以放到苏氨酸 tRNA 合成酶的口袋中。虽然错误的氨基酸与酶的亲和性较低，但这还不足以完全区分相似的氨基酸。这些氨基酸相似性大多可以在Ⅰa 型、Ⅱa 型中找到。错误的氨酰 tRNA 或错误激活的氨基酸需要被清除掉，而这就是通过某些 tRNA 合成酶所拥有的校正机制（editing mechanism）实现的。例如，Ⅰa 型中的亮氨酸、异亮氨酸和缬氨酸 tRNA 合成酶都具有一个保守的被称为 CP1 的校正结构域，但苏氨酸和脯氨酸 tRNA 合成酶中的编校结构域却并不保守。从图 8.9 的异亮氨酸 tRNA 合成酶的结构可以看出，清除错误配对氨基酸的机制，即该酶的编校结构域具有缬氨酸的结合位点和水解酶活性，可以结合缬氨酸并将其水解，而这个结合位点对于正确配对的异亮氨酸而言却太小，因而无法将其水解。

8.3　核糖体

8.3.1　核糖体的组成和功能

核糖体是一个巨大的、由蛋白质和核酸组成的复合体，它是翻译信使 RNA 并合成蛋白质的分子机器。核糖体由一个大亚基和一个小亚基组成，大小亚基间可以聚合和解聚，小亚基负责结合并解码信使 RNA，大亚基负责肽酰转移的过程。与身为蛋白质酶的 DNA 或 RNA 聚合酶相比，核糖体的很多重要功能依赖于其上的 RNA 分子。在细菌中，小亚基（30S）有一个 RNA 分子——16S 核糖体 RNA（rRNA）。S 是 Svedberg unit 简写，是一种在超速离心中对沉降速率的度量单位。小亚基大约由 21 个蛋白质（标记为 S1 至 S21）组成。细菌中的大亚基（50S）有两个 RNA 分子，分别为 5S 和 23S rRNA，还有大约 33 个蛋白质分子（标记为 L1 至 L36）。此外，蛋白质在核糖体中的合成还离不开 mRNA 和 tRNA 分子的参与。

古细菌的核糖体大小和 RNA 分子数与细菌的较为接近，但小亚基和大亚基分别有 28 和 40 个蛋白质。真核生物的核糖体要大得多，小亚基（40S）有一条 18S RNA 和 32 个蛋白质分子；大亚基（60S）有三种 RNA，分别为 5S、5.8S 和 28S，以及大约 46 个蛋白质分子。细菌、古细菌和真核生物核糖体中部分蛋白质是相似的（表 8.3）。

表 8.3 细菌、古细菌和真核生物中各自核糖体的组成

来源	大小	RNA	蛋白质
细菌核糖体	70S	—	—
小亚基	30S	16S	21
大亚基	50S	23S, 5S	33
古细菌核糖体	70S	—	—
小亚基	30S	16S	28
大亚基	50S	23S, 5S	40
真核生物核糖体	80S	—	—
小亚基	40S	18S	32
大亚基	60S	28S, 5.8S, 5S	46

大小两个亚基间通过大量的亚基间相互作用连接在一起（图 8.10，左 B1a~B8）。RNA、蛋白质、镁离子及水分子调控着这些亚基间的相互作用。在发生移位(translocation)时，部分亚基间的相互作用会发生改变。在核糖体上主要有三个 tRNA 分子的结合位点，即 A 位点（amino acid site）、P 位点（peptide site）和 E 位点（exit site）（图 8.11）。另外还有一个额外的 T 位点，是 tRNA 与延伸因子 EF-Tu 和 GTP 的复合体最初的结合位点。

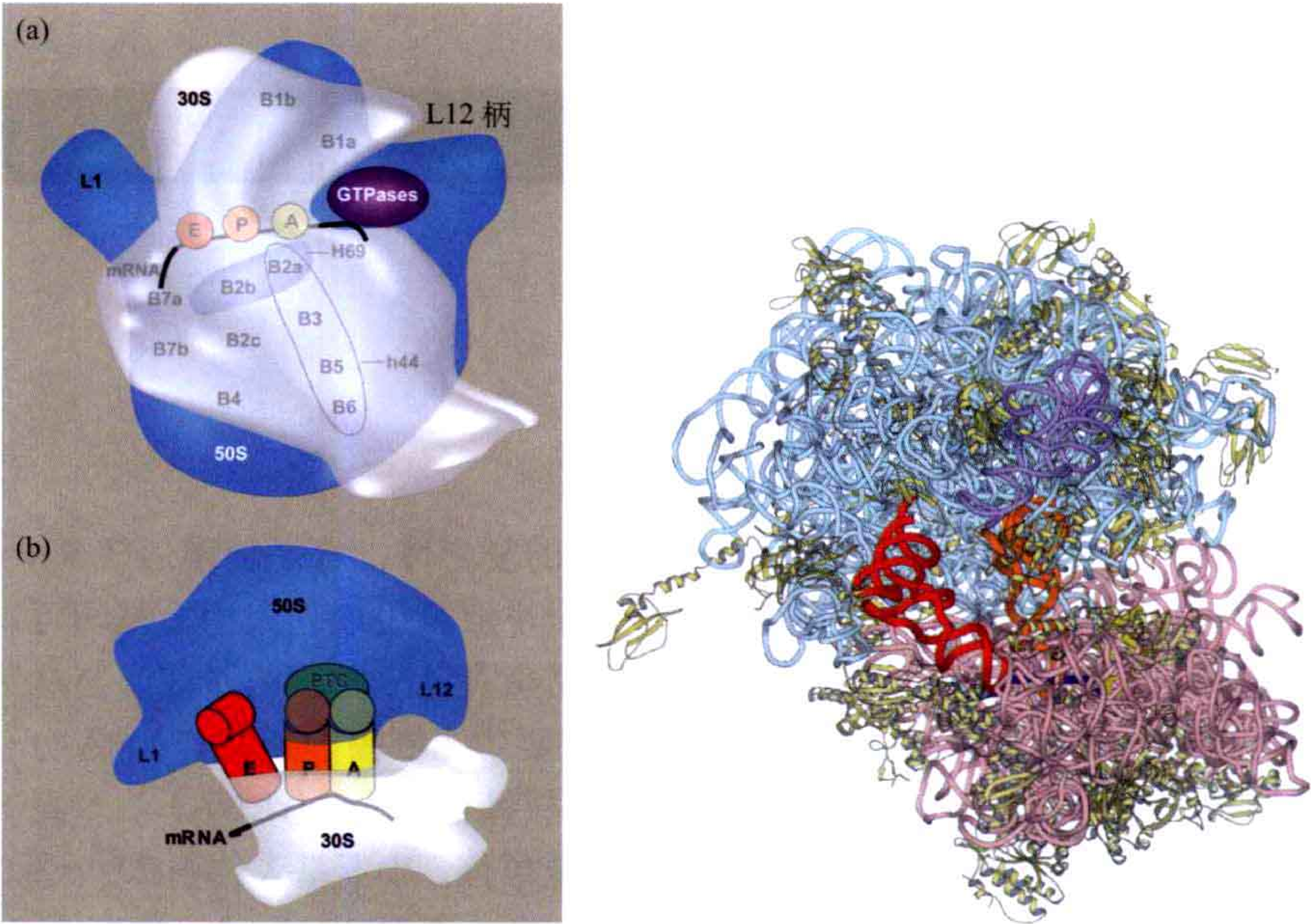


图 8.10 左图：（a）细菌核糖体示意图。50S 大亚基位于后面，30S 小亚基位于前面；核糖体功能位点位于亚基之间，mRNA 结合在小亚基头部和身体之间的颈部；A、P、E 分别表示 tRNA 结合的三个位点；GTPase 表示翻译因子 GTPase 的结合位点；B1~B8 表示亚基间的相互作用位点；h44（30S）和 H69（50S）是两个特殊的 RNA 螺旋，它们在负责 tRNA 解码的 A 位点（即图中 B2a）有一个功能上极为重要的相互作用。（b）核糖体俯视图。能够看到 mRNA、tRNA 结合位点及肽酰转移中心（PTC）。[经 Liljas 授权。原文来自（2006）Deepening ribosomal insights. *ACS Chem Biol* 1: 567-569. Copyright ACS.] 右图：*T. thermophilus* 的 70S 核糖体的晶体结构，分辨率 2.8Å。大亚基的 RNA 分子由淡蓝色和淡紫色表示，小亚基的 RNA 分子由粉色表示；位于 E、P 和 A 三位点的 tRNA 分别用红色、橙色和黄色表示，蓝色表示部分 mRNA 分子；核糖体蛋白质均用黄色表示（PDB：2J00 和 2J01）。

8.3.2　翻译步骤概述

翻译通常可以分为4步进行：起始（initiation）、延伸（elongation）、终止（termination）和再循环（recycling）（图8.11）。在翻译的起始阶段，mRNA分子结合到小亚基上，同时起始tRNA分子结合到P位点上，随后大亚基进一步结合上来。在细菌中，起始由三个起始因子（initiation factor）协助催化，它们是IF1、IF2和IF3。其中IF2具有GTP酶活性。

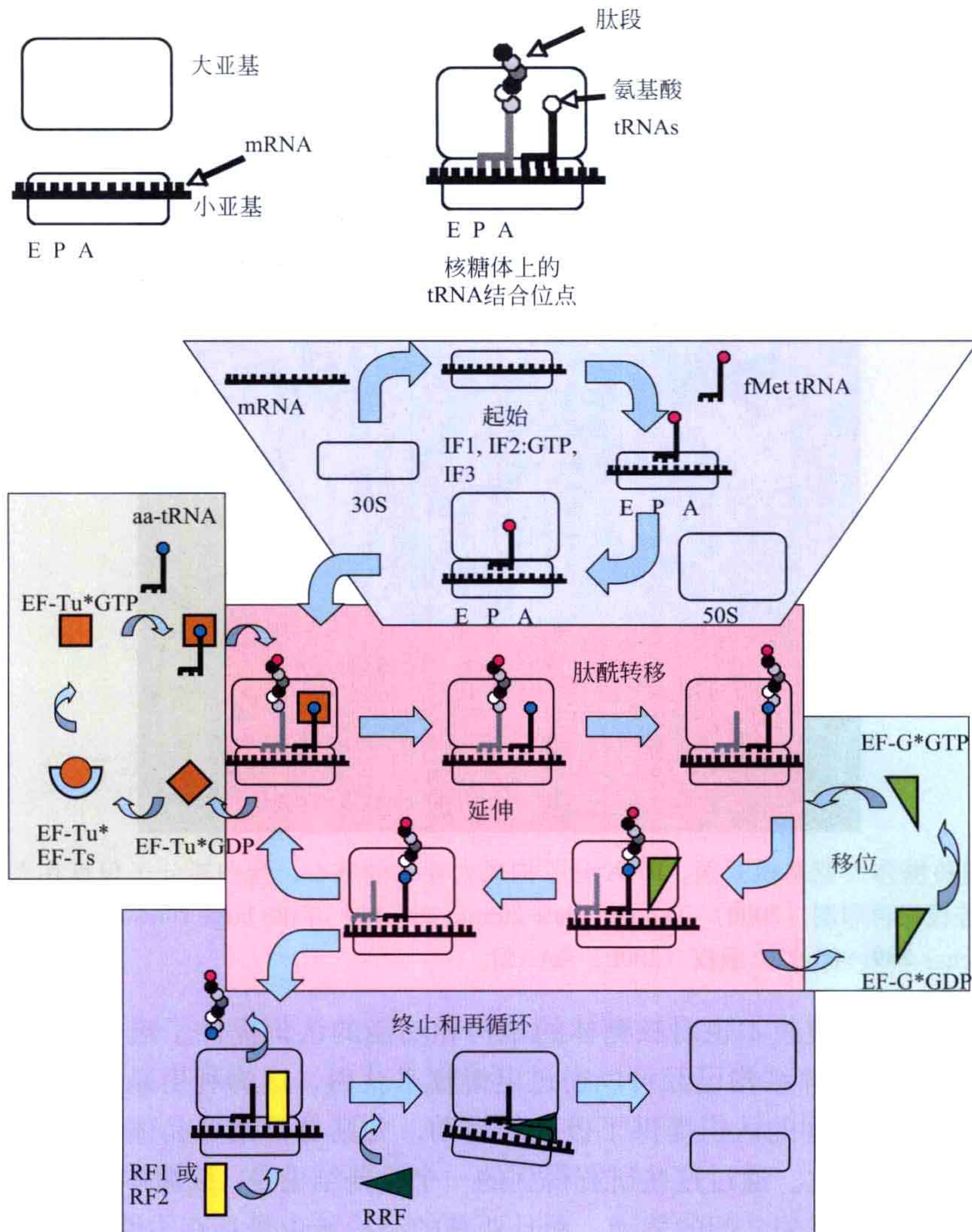


图8.11　细菌中蛋白质合成的主要步骤。

延伸阶段又可以分为三步：氨酰tRNA的解码和结合、肽酰转移及移位。多肽链沿着mRNA移动并随着核糖体翻译的进行而加长。当遇到终止密码子时，多肽链才被释放，随后翻译过程中所有的组件解聚并再循环。这些步骤中的大多数不仅需要核糖体蛋白

和 RNA，还需要各种不同的蛋白质因子的协同作用。

8.3.3 核糖体的结构研究

由于核糖体的大小和复杂性，对于整个核糖体或核糖体大小亚基的研究非常困难，对核糖体的结构研究更是一个巨大的挑战。迄今已经有来自两个物种的完整核糖体结构得到了解析，它们分别为大肠杆菌（*E. coli*）（3.5Å）和嗜热菌（*Thermus thermophilus*）（2.8Å）。而部分核糖体活动可以通过对分离的亚基的研究获知，因此单独亚基的结构研究也具有重要意义。例如，来自古细菌 *Haloarcula marismortui*（2.4Å）（图 8.12）和细菌 *Deinococcus radiodurans*（3.1Å）的 50S 亚基的晶体结构也已经分别得到了解析。

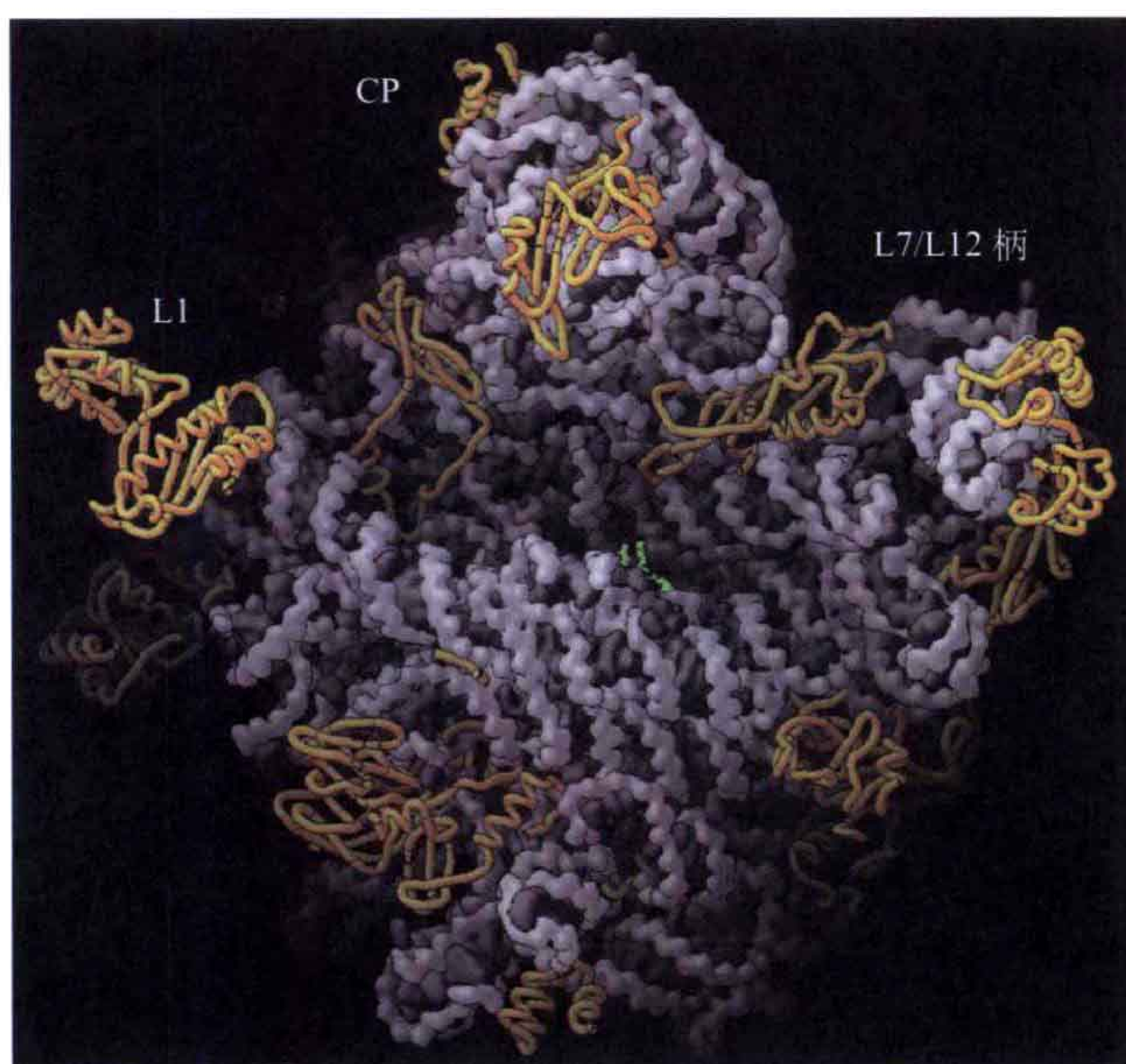


图 8.12 细菌核糖体大亚基模式图。RNA 分子构成大亚基的核心，蛋白质分子包裹在大亚基的表面 [经 Bann 等授权再印制（2000） The complete atomic structure of the large ribosomal subunit at 2.4Å resolution. *Science* **289:** 905-920. 版权（2000）AAAS]。

除了晶体学，电镜技术也对核糖体的结构和功能的认识提供了进一步的帮助。高于 10Å 分辨率的核糖体结构已经可以通过电镜技术获得，这为利用晶体学方法无法观测到的蛋白质合成过程的认识提供了极大的帮助，尤其是再结合晶体学结构提供的细节上的支持时更为如此。通过这些研究得到的一个重要结论是：核糖体是具有柔性的，核糖体亚基能够影响互相之间的移动，而且亚基的结构域也是具有柔性的。

8.3.3.1 大亚基的结构

从侧面观察时大亚基呈皇冠状，从顶面观察时呈半球形。三个延伸出来的部分从左至右依次称为 L1 柄（L1 stalk）、中心突起（central protuberance）（5S RNA 定位于

此）和L12柄（L12 stalk）。两边的突出部分主要由蛋白质组成，并包含明显的柔性区。

大亚基的核心是由23S RNA组成的，23S RNA可划分为6个结构域，这6个结构域相互紧密缠绕使大亚基形成一个稳定的结构。23S RNA的二级结构由大约100条双链螺旋组成，此外还有许多三级结构相互作用。

蛋白质主要位于核糖体的表面，这些蛋白质大多有着不同寻常的结构特征（图8.13），有的蛋白质在核糖体表面的部分呈球状结构，同时具有延伸部分可以深入亚基内部与RNA进行相互作用；而有的核糖体蛋白完全呈现出能够与RNA作用的延伸结构，这些延伸结构很可能对核糖体的聚集和稳定有着极其重要的作用。

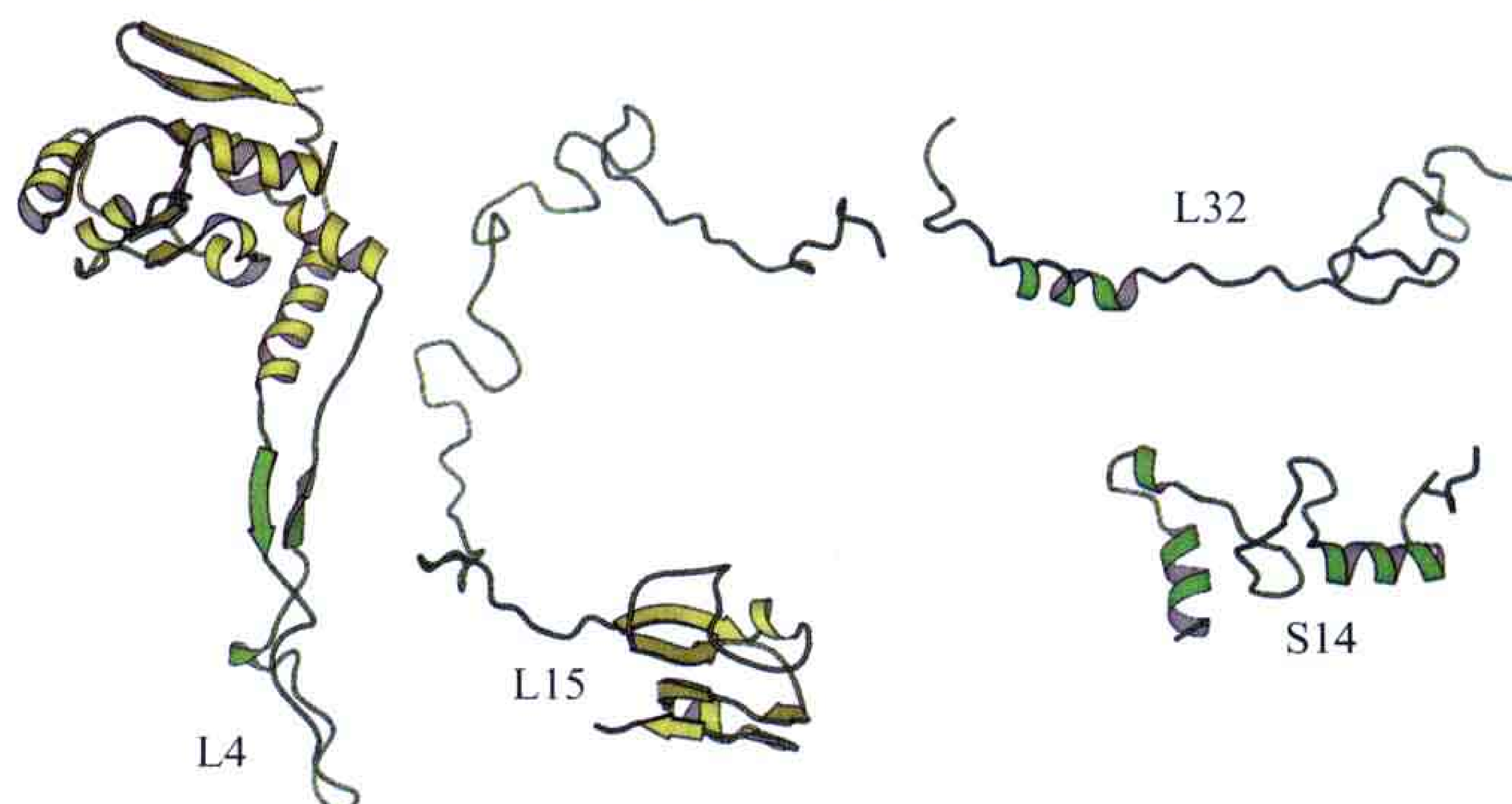


图 8.13　一些表现出不寻常构象的核糖体蛋白结构。大多数核糖体蛋白拥有延伸的末端，结构域之间由不同长度的序列相连。蛋白质中可能只在与核糖体结合后才发生折叠的部分由绿色表示。这些蛋白质均来自 *Thermus thermophilus* 核糖体（PDB：2J00 2J01）。

大亚基的主要功能是催化肽键的形成，而肽键的形成是由肽酰转移中心（peptidyl transfer center，PTC）完成的。PTC主要由23S RNA组成，但在细菌核糖体中蛋白质L27的定位与PTC的RNA极其接近。

增长中的多肽链从大亚基的一个开口伸出（图8.14）。这个出口开始于大亚基的PTC，结束于大亚基的外表面。这条大约100Å长的通道主要由RNA和几种蛋白质构成。

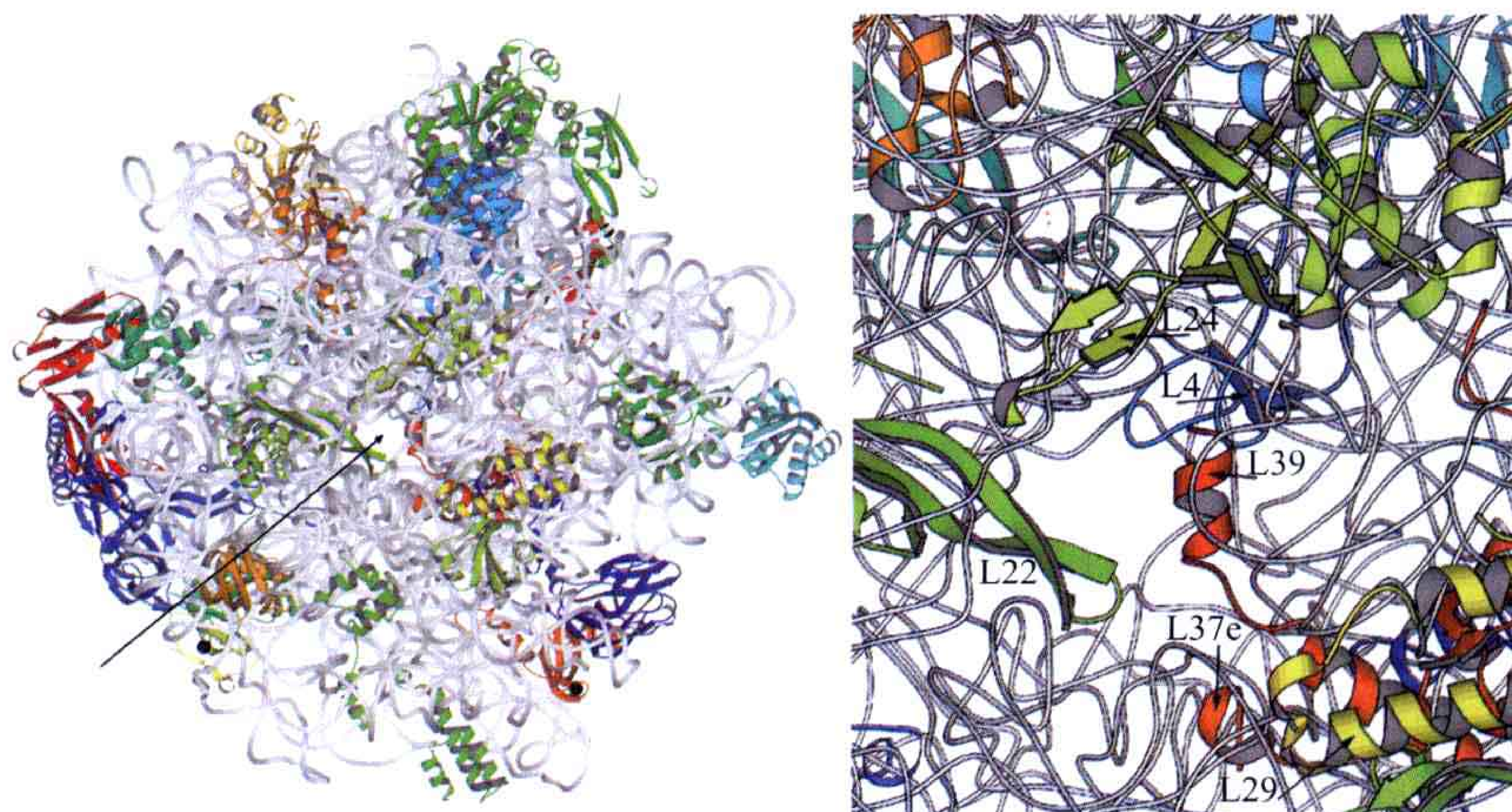


图 8.14　左：核糖体大亚基俯视图。箭头指示出了多肽段的出口通道，卷曲部分为RNA。右：多肽段出口通道放大图。蛋白质L4和L22的延伸部分是该通道的重要组成部分。

8.3.3.2 小亚基的结构

除了 *E. coli* 和 *T. thermophilus* 的完整的核糖体结构得到解析外，*T. thermophilus* 的小亚基结构也得到了解析。与大亚基类似，小亚基的核心部分也是由 RNA 组成的。与大亚基的结构域相比，小亚基的 4 个结构域被清晰地分割开来，从而使小亚基在相对指向上具有更多的可变性。如图 8.15 所示，红色部分为小亚基的体部（body），由 16S RNA 5′ 端的大约 1/3 的部分组成；绿色部分为小亚基的平台（platform），它是 16S RNA 的核心部分；黄色部分为小亚基的头部（head），头部长有一个喙（beak）；蓝色部分为 3′ 端，它形成一条垂直的螺旋（h44）。蛋白质主要位于小亚基的表面，并部分具有连接和稳定小亚基 RNA 螺旋间相互作用的功能。

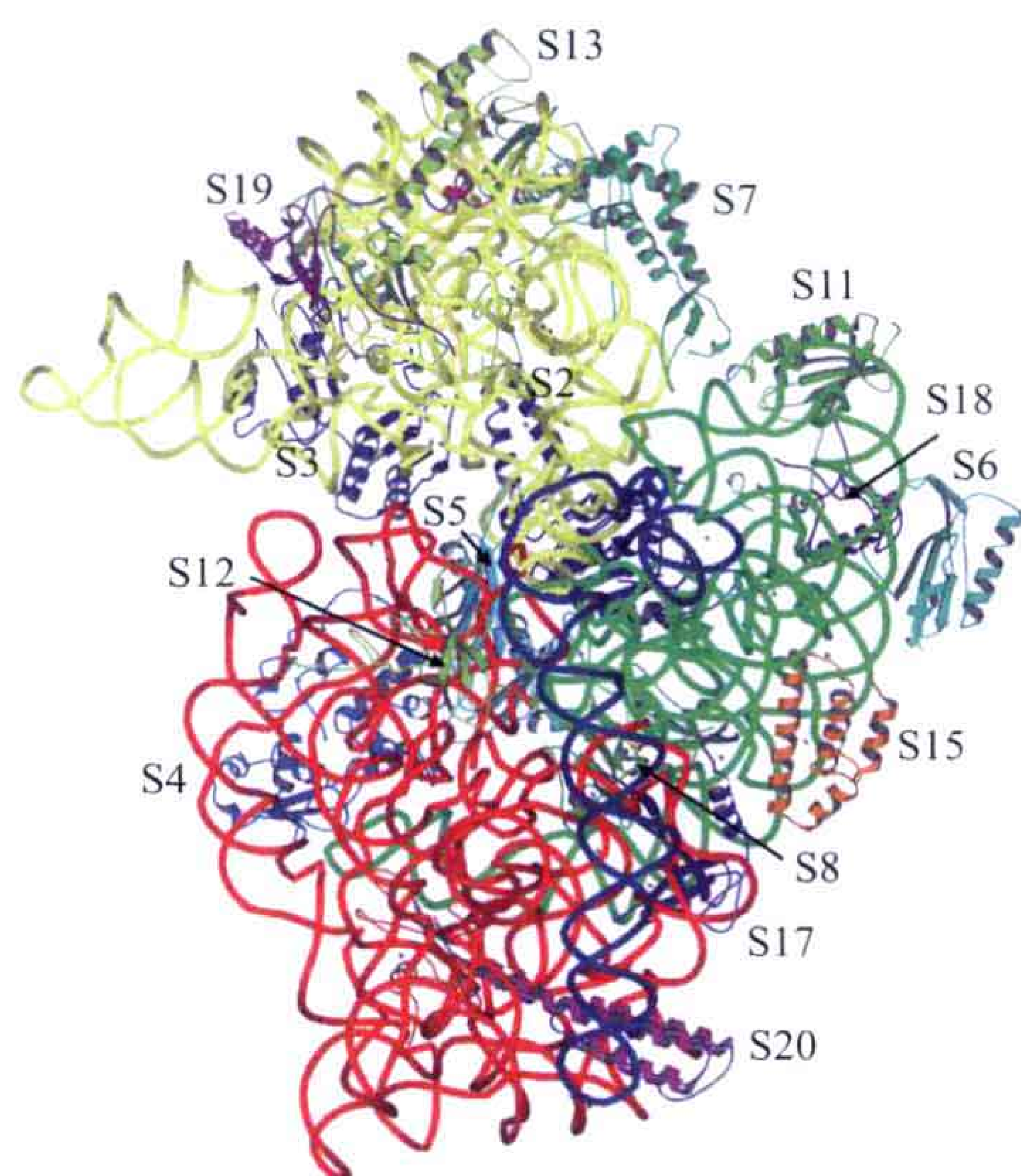


图 8.15 核糖体小亚基。RNA 结构域从 5′ 端至 3′ 端分别用红色、绿色、黄色和蓝色的飘带表示，同时标注了小亚基的大部分蛋白质组分。

小亚基能够结合 mRNA，它的主要功能是参与 mRNA 的解码。小亚基还部分参与了核糖体的 A 位点、P 位点和 E 位点的构成。通过大量对核糖体的生化、晶体学及电镜试验，这三个位点的定义与划分已经越来越清晰。其中对于 A 位点的最清晰的研究便是来自于对核糖体小亚基与 tRNA 片段和 mRNA 片段的复合体的研究。

8.4 翻译的起始

在细菌中，三种不同的蛋白质因子 IF1、IF2 和 IF3 参与催化了翻译的起始。这些蛋白质因子的功能是帮助 mRNA、核糖体亚基及起始 ^{fMet}tRNA 三者形成正确的复合体。

IF1 是一个广泛保守的带有 OB 折叠（OB-fold）的小蛋白，它能够结合到 30S 亚基 A 位点的解码区，IF1 与该位点的结合很有可能起到了阻止起始 tRNA 结合到 A 位点的作用，而只允许它结合在 P 位点。此外，IF1 还参与了稳定 mRNA 和起始 tRNA 间的密码子 - 反密码子相互作用。

IF2 属于具有 GTP 酶活性的 G 蛋白家族的 GDP/GTP 结合蛋白（见 5.3 节）。IF2 的主要作用是催化起始时大亚基与小亚基的结合。

IF3 结合在核糖体小亚基上，并能防止大亚基与小亚基间过早地发生结合。IF3 可明显分为 N 端和 C 端两个结构域，而两个结构域的结构已经分别得到了解析。N 端结

构域的结合位点已知为 E 位点 30S 亚基的一部分，而 C 端结构域结合在小亚基与大亚基相互作用的界面上，从而起到阻止亚基间聚合的作用。

起始密码子 AUG 同时还编码了甲硫氨酸。而 mRNA 5′ 端富含嘌呤的区域的 SD 序列（Shine-Dalgarno）通过与 16S RNA 3′ 端的一段互补序列的相互作用帮助避免了从任何一个随机的 AUG 都能发生起始的情况。在细菌中，起始氨基酸是甲酰甲硫氨酸（formyl-methionine，fMet）。mRNA 与起始的 $tRNA^{fMet}$ 在 IF1 和 IF3 的帮助下能够结合到核糖体小亚基上。$fMet\text{-}tRNA^{fMet}$ 结合在核糖体 P 位点的 AUG 密码子上，然后大亚基在 IF2 和 GTP 的帮助下结合上来，形成完整的起始复合体。

在真核生物中，翻译的起始过程比原核生物复杂得多，有更多的起始因子参与了反应。在真核生物起始中发现了 2 个 G 蛋白，其中 eIF2 的作用是帮助起始 tRNA 结合到核糖体上；而另一个 G 蛋白，即 eIF5 是细菌中 IF2 的同源蛋白，同样起着连接大小亚基的作用。该蛋白质有 4 个结构域，其中的 N 端结构域是一个 G 结构域，与其他的 G 蛋白功能相似（图 8.16）。前两个结构域与 EF-Tu 中相应的结构域类似。

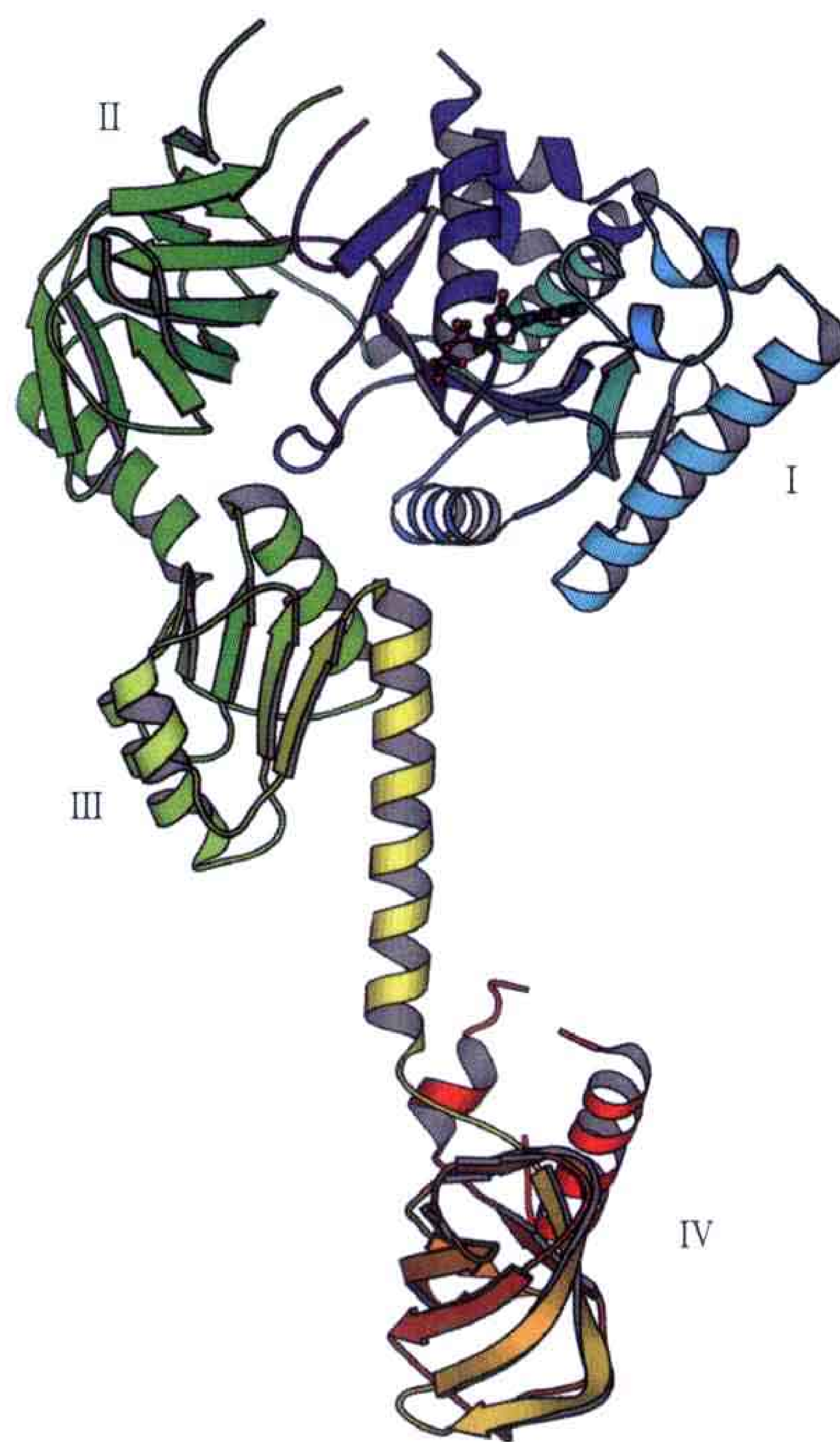


图 8.16 来自嗜热自养甲烷杆菌（*Methanobacterium thermoautotrophicum*）的 eIF5B 的 GDP 结合形态。eIF5B 相应于细菌中的 IF2。结构域 I 是 G 结构域（PDB：1G7S），结构域 IV 与起始 tRNA 的 fMet 部分相互作用。

8.5 延伸

在细菌的延伸中，主要有两个蛋白质因子参与了催化 tRNA 与核糖体的结合和在核糖体上的移位（表 8.4），它们是 EF-Tu 和 EF-G。这两个蛋白质都是 G 蛋白，而且通常它们在一次延伸反应中只催化水解一个 GTP 分子生成 GDP。在延伸中 tRNA 在 EF-Tu 和 GTP 形成的复合体的帮助下结合到 A 位点。在肽酰转移完成后，EF-G 负责催化移位反应。在移位反应中，原来位于 A 位点的肽酰 tRNA 进入 P 位点，而去肽酰化的 tRNA 会进入 E 位点。在这个过程中 mRNA 也会同时向前移位以使新的密码子进入 A 位点。在进一步讨论延伸反应前，我们先来探讨一下各延伸因子的结构。

表 8.4 细菌和真核生物中的延伸因子

蛋白质	功能
EF1A（EF-Tu）	帮助连接氨酰 tRNA 与核糖体的 G 蛋白
EF1B（EF-Ts）	EF1A/EF-Tu 的核苷酸转换因子。细菌与真核中的该蛋白质是不相关的。真核的 eEF1B 有两个亚基，即 α 和 γ，其中 α 是活性的转换因子，而 γ 与谷胱甘肽转移酶非常相似
EF2(EF-G)	帮助肽酰 tRNA 从 A 位点移位至 P 位点的 G 蛋白
SelB	相应于硒代半胱氨酸对应的 EF-Tu，硒代半胱氨酸是一种稀有的第 21 种氨基酸，通过特殊的 tRNA 在延伸过程中连接到多肽上
Lep A	同样是一个 GTP 酶，可能具有逆向移位的作用
Tet M，O，S 等	四环素抗性因子，与 EF-G 高度同源

注：括号中是细菌中对应因子的常用名。

8.5.1 延伸因子 EF-Tu

延伸因子 EF-Tu 负责将氨酰 tRNA 运输至核糖体。EF-Tu 是一个 G 蛋白，依赖于 GDP/GTP 的开关机制可以帮助控制 tRNA 上反密码子和核糖体上 mRNA 密码子间正确的配对。EF-Tu 在结构和生化方面都已经得到了广泛的研究，因此我们在这里将对它的活性如何得以调控展开详细的讨论。

单独的 EF-Tu 本身的 GTP 酶活性是很弱的，但是就像 GAP 蛋白可以激发其他 G 蛋白的活性一样，EF-Tu 的活性可以由核糖体激发。具体的机制尚不清楚，但很可能核糖体的一部分参与了激活过程。当 tRNA 的反密码子与密码子配对时，就诱导了 GTP 水解反应的发生。当 GTP 被水解后，EF-Tu 的构象会发生变化并被从核糖体上释放出来。

EF-Tu 由三个结构域构成。N 端结构域主要由平行 β 折叠片组成，并与其他 G 蛋白有着相同的拓扑结构（图 8.17）。其他的两个结构域都由反平行 β 桶组成。

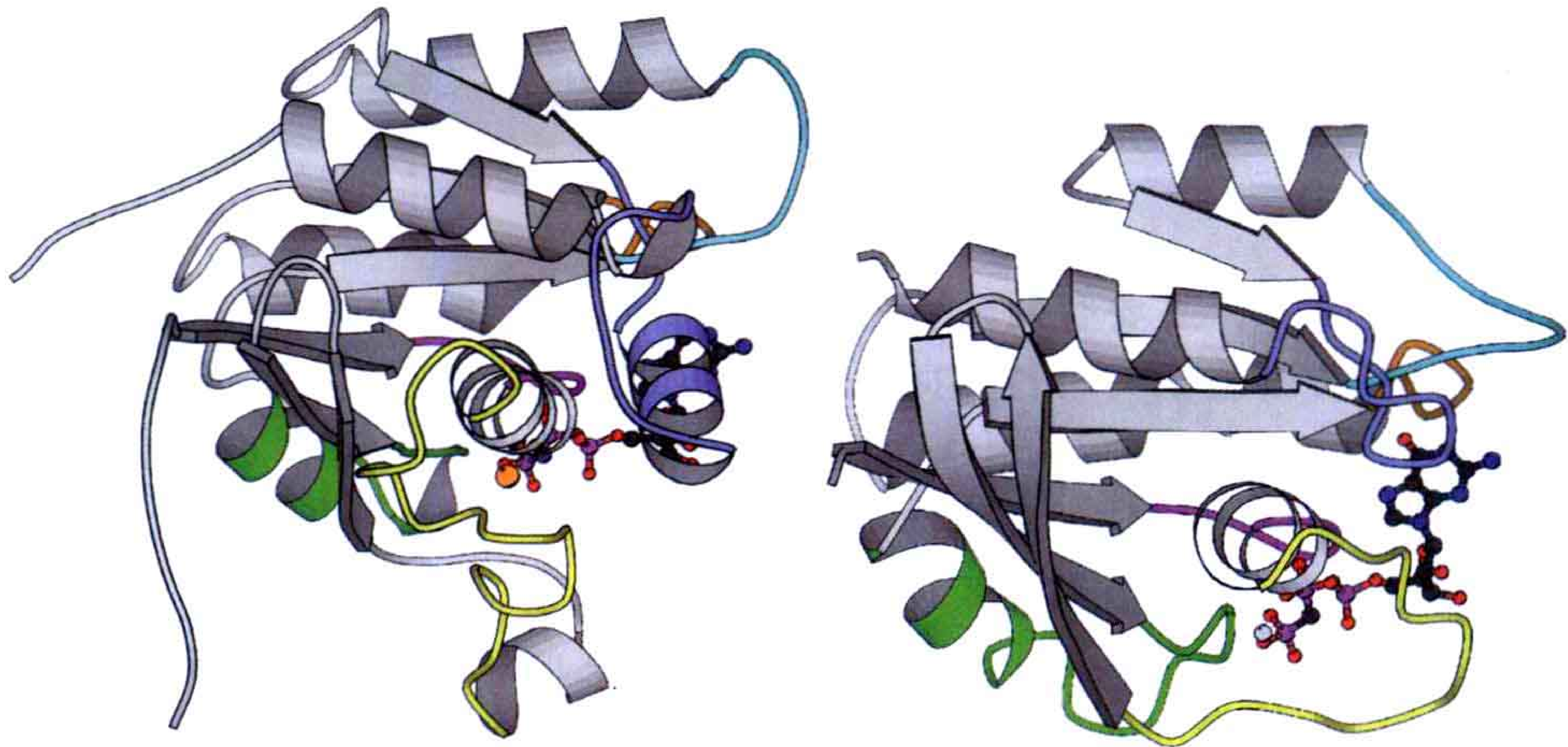


图 8.17 EF-Tu（左，PDB：1EFT）与信号蛋白 Ras（右，PDB：121P）的 N 端结构域 1 的比较。紫色：G1，P-loop；黄色：G2，开关Ⅰ；绿色：G3，开关Ⅱ；蓝绿色：G4；淡蓝色：G5。

EF-Tu 的 GTP 和 GDP 形式在构象上有着很大的差别。在 GTP 形式中，三个结构域紧密地折叠在一起；而在 GDP 形式中，两个 β 桶结构域远离 G 结构域而使 EF-Tu 形成一个更加开放的构象（图 8.18）。蛋白质的某些区域在两种构象中的差别可达 40Å。结构域 2 和 3 在两种形式中保持了它们的相对走向。

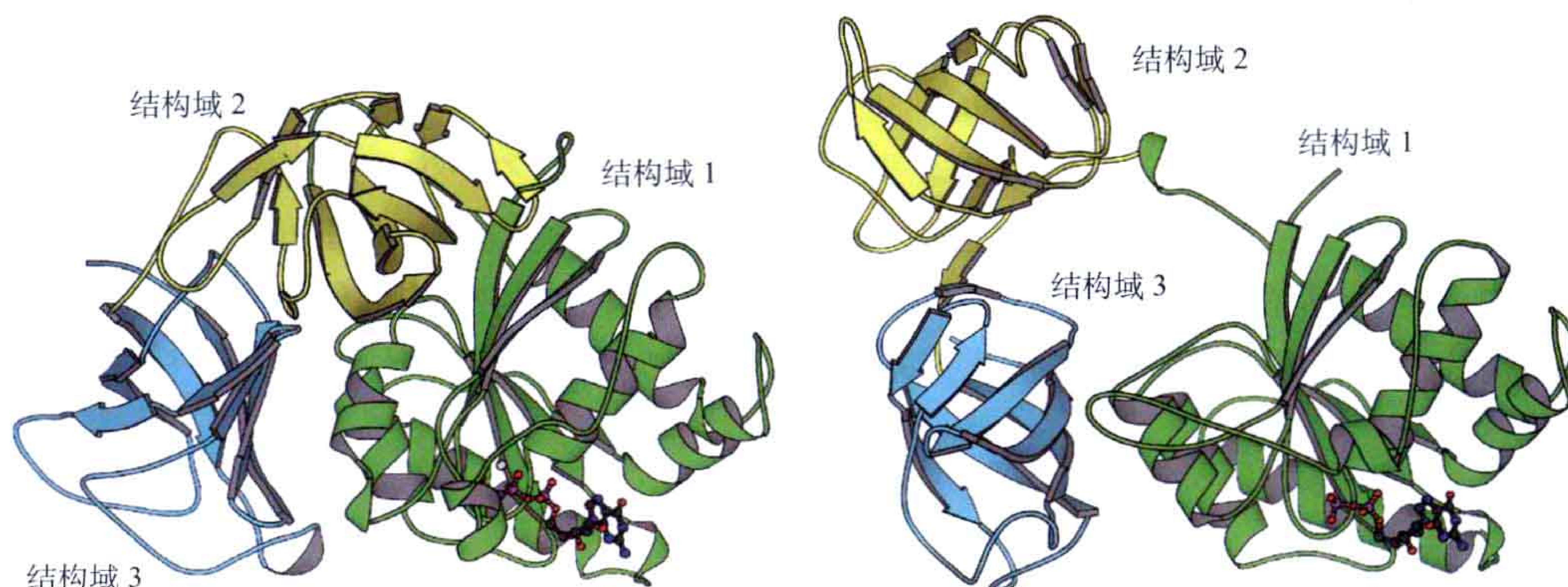


图 8.18　*T. aquaticus* EF-Tu 的 GTP 构象（左，PDB：1EFT）和 *E.coli* EF-Tu 的 GDP 构象（右，PDB：1TUI）。

GTP 与 EF-Tu 的结合方式与其他 G 蛋白类似：P-loop 能够结合 α 和 β 磷酸，一个镁离子用于调控 GDP 形式中的 β 磷酸，以及 GTP 形式中的 β 和 γ 磷酸（图 8.19）。但与 Ras 和三聚 G 蛋白（trimeric G-protein）相比，开关区（switch region）表现出了不同的构象变化。

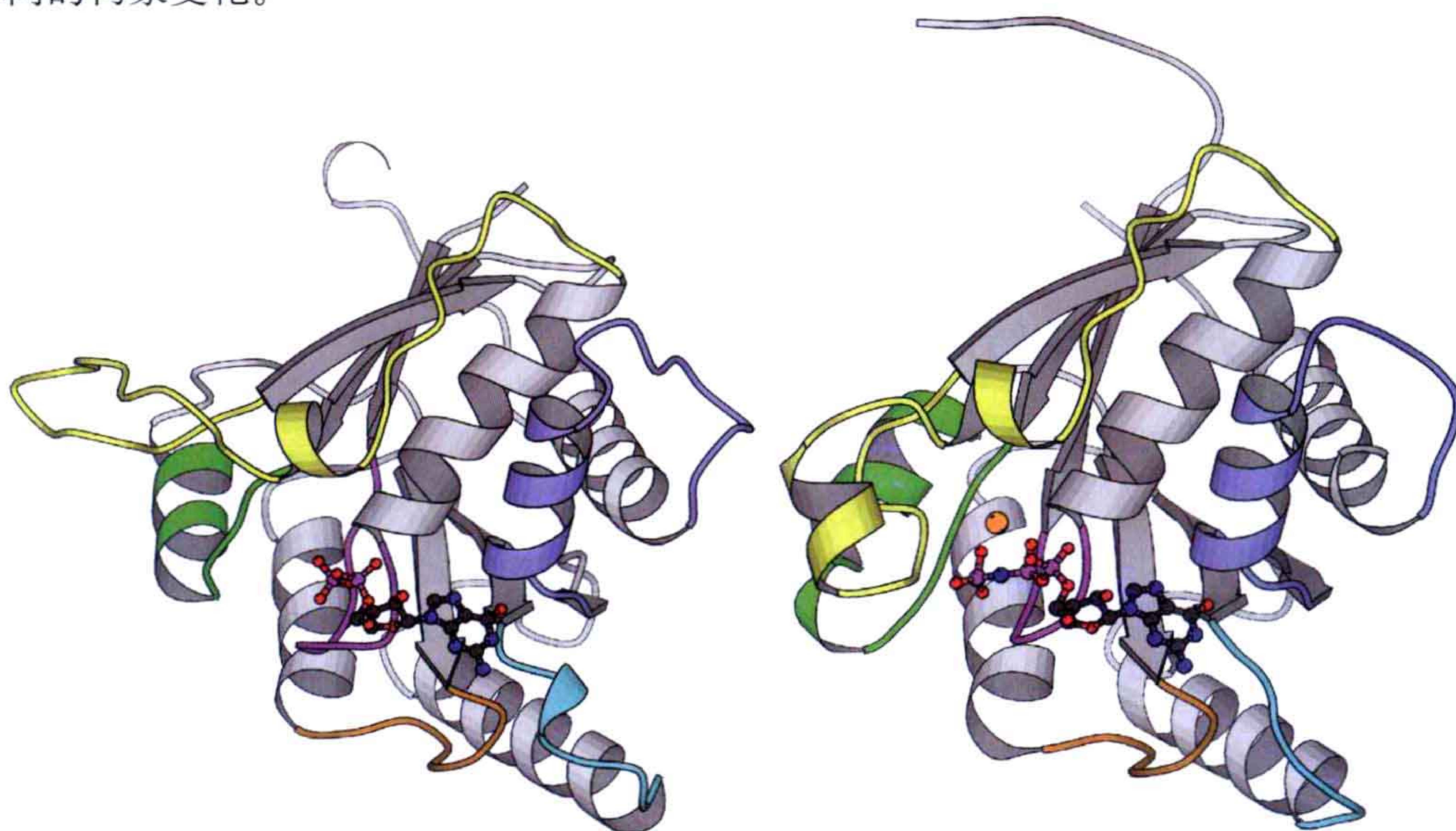


图 8.19　GDP、GTP 与 EF-Tu 的结合。loop 的颜色如图 8.17 所示。在 GDP 构象中左边延伸出来的 loop（开关 I）在 GTP 构象中变成了一段能与 γ 磷酸相互作用的螺旋。

与 Ras 中的情况类似，GDP 分子的磷酸能够与 P-loop 相互作用，但与其他的

loop几乎没有相互作用。在有GTP存在时，开关区Ⅰ和Ⅱ的构象会发生改变（图8.20）。在GDP形式中，开关Ⅱ的螺旋部分由残基85~94组成；而当结合GTP时，这个螺旋结构之前的loop必须要发生构象变化，因为残基83的羰基氧此时与GTP的γ磷酸过于接近。因此在GTP形式中该肽键与GDP形式相比发生了一个翻转，从而可以使残基84的氨基氮与磷酸间形成氢键，同时这会使螺旋部分“解旋”，使loop延伸至残基88，而新的螺旋（残基89~96）也具有一个不同的走向。在GTP形式中螺旋部分与结构域3紧密相连。GTP改变了开关Ⅱ的构象，使之产生了一个与GTP形式中的其他结构域相互作用所必需的表面，并激活与tRNA的结合。在其他的GTP酶（如Ras）中，与酶活性紧密相关的是一个谷氨酰胺，而在能与核糖体结合的GTP酶中该氨基酸是一个组氨酸残基。这些残基对在水解时在γ磷酸处放置一个水分子是极为重要的。

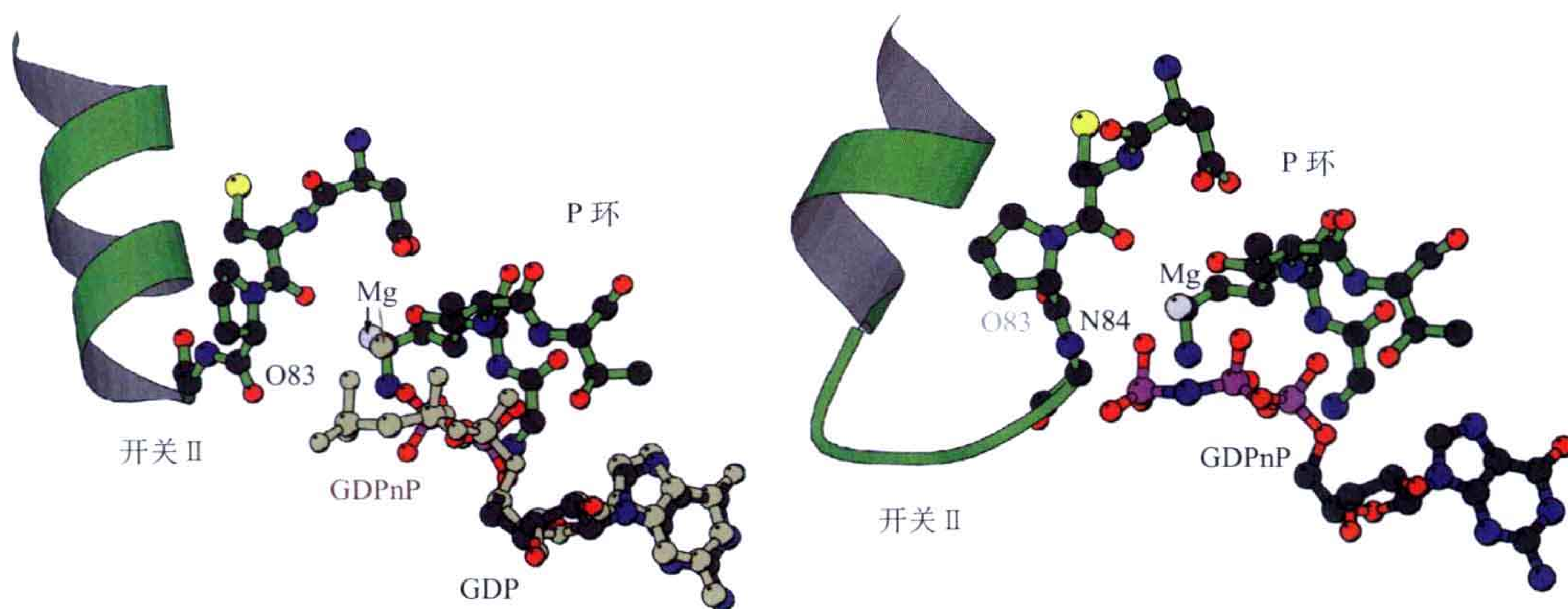


图8.20 GDP/GTP结合部位细节图。左：带有一分子GTP的GDP构象，展示了与83位的羰基氧有紧密的相互作用；右：GTP构象，螺旋结构具有新的位置和走向。

8.5.1.1 与tRNA的结合

转运RNA与EF-Tu的结合同tRNA与氨酰tRNA合成酶的结合完全不同。因为一个EF-Tu能够结合所有不同的tRNA分子，它只需识别tRNA分子上的保守位点。此外，EF-Tu还必须能区分携带氨基酸和不携带氨基酸的tRNA。虽然携带一个氨基酸并没有很大地改变tRNA的表面，但EF-Tu对未携带氨基酸的tRNA的亲和性非常低。而对携带正确和不正确的氨基酸tRNA的结合也是有差别的，在某些情况下这能增加tRNA水平的特异性。

从整个复合体来看，Tu的所有三个结构域都与tRNA分子发生相互作用。只有一小部分tRNA分子，主要来自结构末端和保守的CCA序列，参与了与Tu的相互作用。反密码子臂不参与与Tu的相互作用（图8.21）。

tRNA的3′端与结构域1和2交界面处的氨基酸相互作用。tRNA的5′端结合在所有三个结构域的交界面，接受臂的一个侧面与结构域3相互作用。对于一个携带了氨

基酸的 tRNA 的特异性结合是通过氨基酸和连接氨基酸到 3′ 核苷酸核糖的酯键间的相互作用实现的。

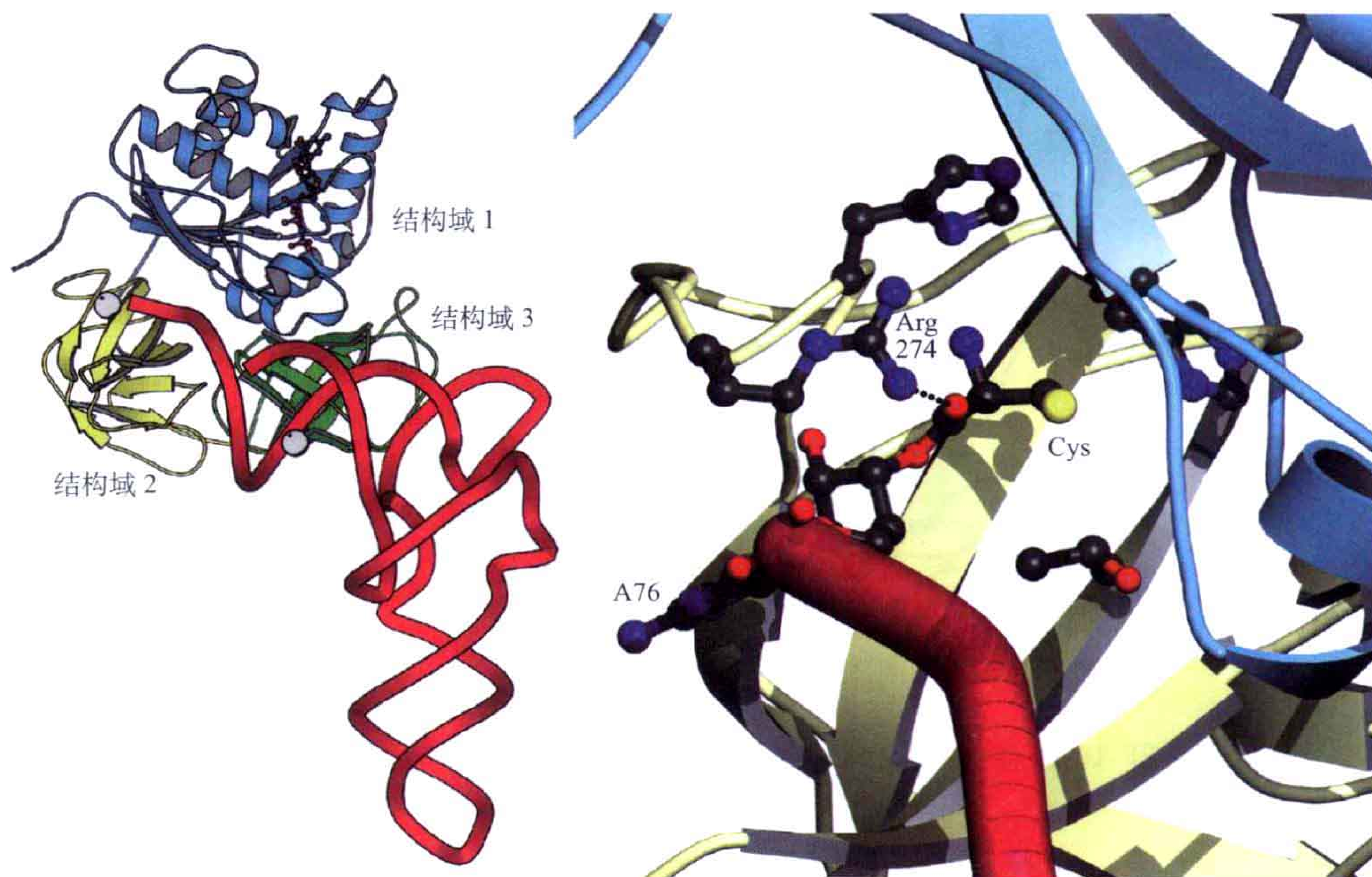


图 8.21　左：EF-Tu 与 Cys-tRNA 的复合体。右：3′ 端（A76）与 tRNA 上的半胱氨酸间的相互作用。274 位精氨酸（Arg274）的胍基与半胱氨酸羰基氧通过氢键相互作用。其他来自于结构域 1 和 2 的侧链形成一个将半胱氨酸紧密包裹在内的牢笼结构。

8.5.1.2　延伸因子 Ts——交换因子

延伸因子 Ts（EF-Ts）是 EF-Tu 的鸟嘌呤核苷酸交换因子（G nucleotide exchange factor，GEF）。该蛋白质不如 EF-Tu 保守，EF-Tu 与 EF-Ts 复合体的晶体结构显示，来自 *T. thermophilus* 的 EF-Ts 为二聚体，而来自 *E.coli* 的 EF-Ts 为单体，但二者执行的功能完全相同。*T. thermophilus* 的 EF-Ts 由三股反平行 β 折叠片组成，其中两股之间还夹着一对螺旋结构。大肠杆菌的 EF-Ts 要长得多，但仍有两个与 *T. thermophilus* 的单体极为相似的区域。这表明大肠杆菌的 EF-Ts 可能通过基因的复制和融合演化而来。

在复合体结构中能看到，EF-Ts 的一个结构域或亚基结合到 EF-Tu 的结构域 1 上，另一个结合在结构域 3 上。EF-Tu 具有同 GDP 相似的构象（图 8.22）。

EF-Tu 结构中的三个变化解释了 EF-Ts 是如何作为一个核苷酸交换因子行使功能的。在复合体结构中，开关 II 的螺旋（螺旋 B）被移开，导致镁离子损失一个配体。另一个变化是 Ts 的 82 位苯丙氨酸插入 Tu 的一个口袋中，直接引起 P-loop 的移位。原来结合到 β 和 γ 磷酸上的 P-loop 的 24 位赖氨酸现在变成与开关 II 的 81 位天冬氨酸相互作用（图 8.23）。另外，残基 20 和 21 间的肽键翻转，也会造成原来结合 GDP 上 β 磷酸

的 N21 被排斥核苷酸的 O20 所替代。

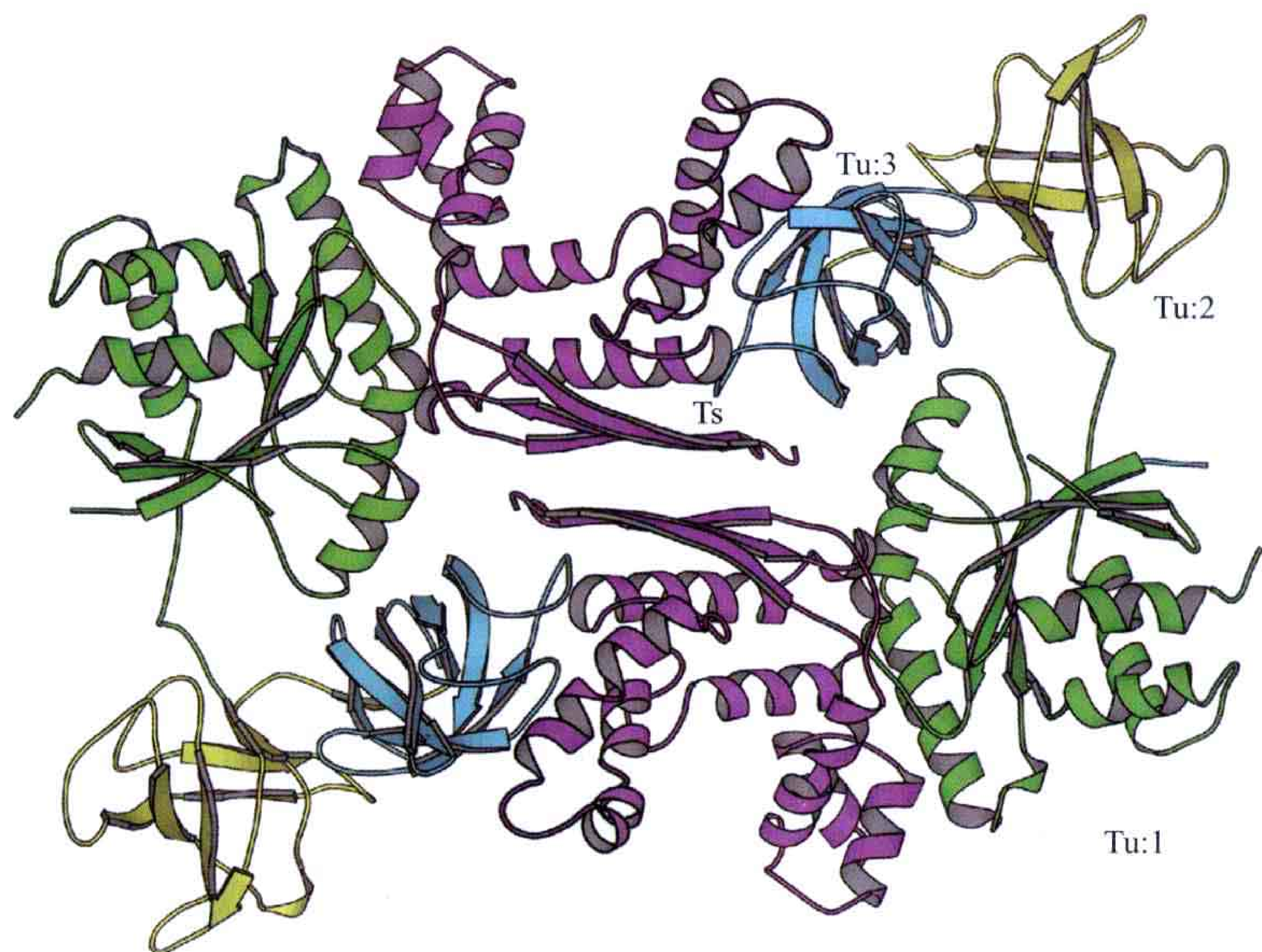


图 8.22　大肠杆菌 EF-Tu-EF-Ts 的复合体结构（PDB：1EFU）。

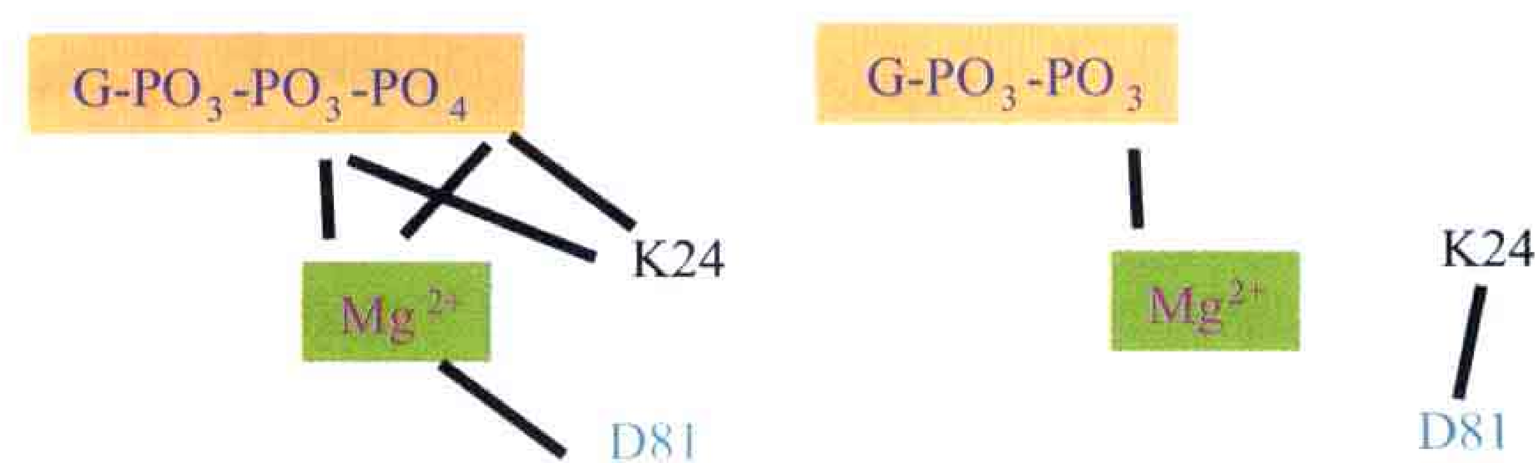


图 8.23　EF-Tu 和 GTP 间（左）及 EF-Tu* EF-Ts 和 GDP 间（右）的一些重要相互作用。EF-Ts 的结合会使 Asp81 与 P-loop 的 K24 相互作用。形成复合体后对镁离子和核苷酸的结合都会变弱，导致核苷酸交换的发生。

8.5.2　延伸因子 G

延伸因子 G（elongation factor G，EF-G，真核生物中的 EF2）催化去氨酰化 tRNA 在核糖体上从 P 位点到 E 位点及肽酰 tRNA 从 A 位点到 P 位点的移位。同时 mRNA 沿核糖体移动并使一个新的密码子暴露在 A 位点。

EF-G 是由 5 个结构域组成的长条形分子（图 8.24）。N 端结构域是一个与其他 G 蛋白有着相同拓扑结构的 G-domain。在 G-domain 的最后一股 β 束前还有一个由反平行 β 片层组成的亚结构域（G′）。结构域 Ⅱ 是所有核糖体相关的 GTP 酶（如 IF2 和 EF-Tu）都含有的一个反平行 β 桶结构。其他的三个结构域都是由一边带有螺旋的反平行 β 折叠片构成的。结构域Ⅲ和 Ⅴ 与核糖体蛋白 S6 及很多 RNA 结合蛋白有相同的拓扑结构。EF-G 分子与 EF-Tu-tRNA 复合体的结构非常相似（可比较图 8.24 和图 8.21）。结构域Ⅲ、Ⅳ和 Ⅴ 占据着类似于一个 tRNA 分子所占据的空间。

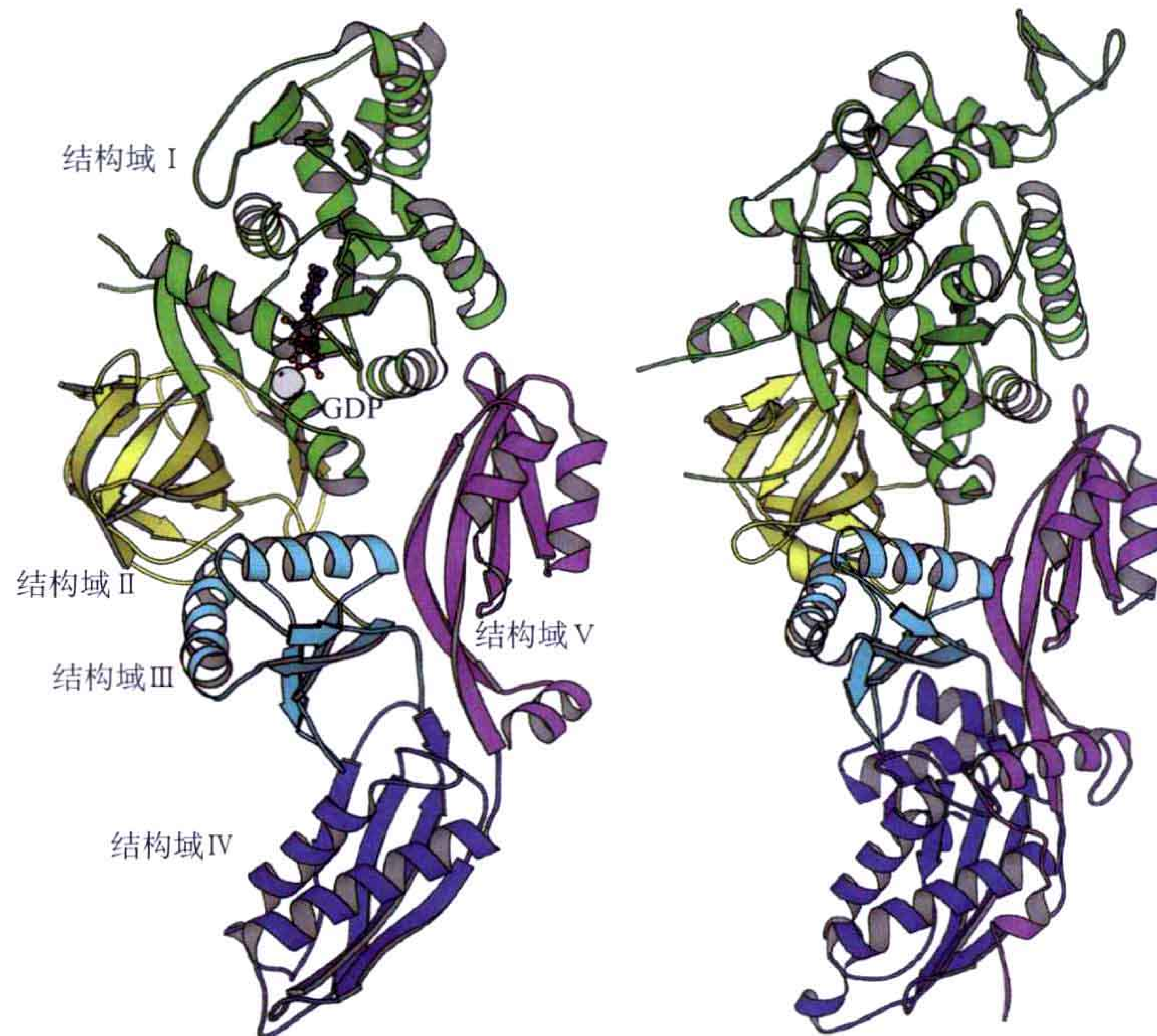


图 8.24 *Thermus thermophilus* 的 EF-G（左，PDB：1FNM）和酵母的 EF-2（右，PDB：1N0V）示意图。

EF-G 不需要核苷酸交换因子，它对核苷酸的亲和性比 EF-Tu 的要低得多，交换可以自发的发生。在 EF-Tu 中，P-loop 的 K24 和开关 II 的 D81 可以稳定能与磷酸相互作用的 GDP 和镁离子。当 EF-Tu 与 EF-Ts 相互作用时，赖氨酸和天冬氨酸会变为与自身相互作用（图 8.25），对于 EF-G 也是如此。在 EF-G 中，是 K25 与开关 II 的 T84 相互作用（图 8.25）。当 EF-G 结合到核糖体上时，这些残基可能会分别与磷酸和镁离子相互作用，而当 EF-G 在溶液中时不会发生这种相互作用。金属离子和核苷酸与 EF-G 的结合都不太紧密，从而使 GDP 交换为 GTP。

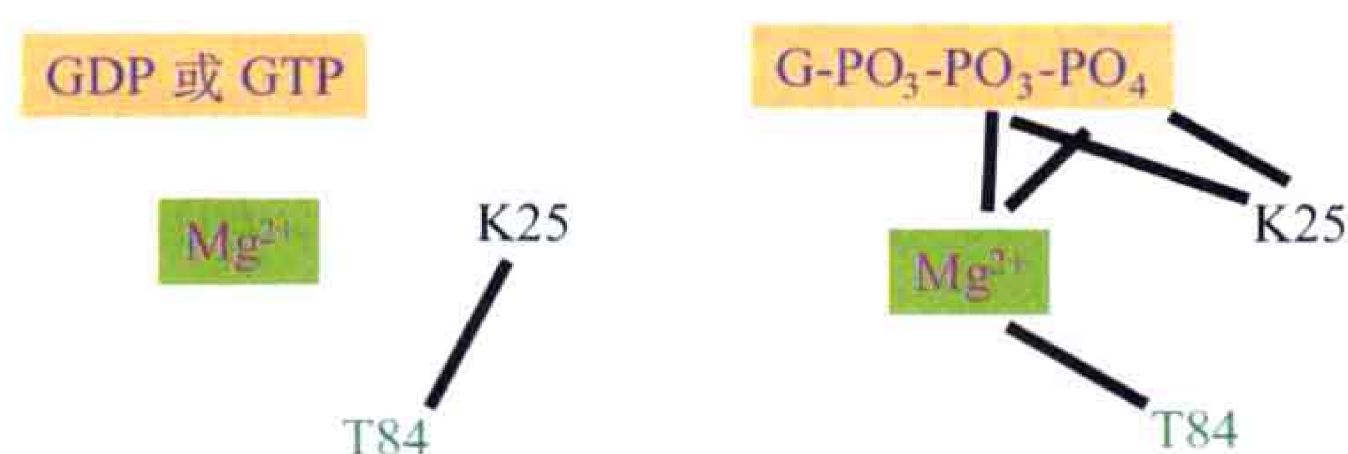


图 8.25 EF-G 的构象使其可以在离开核糖体时自发地发生交换反应。P-loop 的赖氨酸通常与开关 II 的 T84 相互作用（左）。当在核糖体上时，GTP 结合得更为紧密，这可能是由赖氨酸和苏氨酸残基分别与磷酸和镁离子的相互作用造成的（右）。

8.5.3 tRNA 的结合与解码——延伸的功能性步骤

在延伸的第一步中，氨酰 tRNA 及 EF-Tu*GTP 组成的三聚复合体需要先在 A 位点进行密码子的检测。通过冷冻电镜发现，氨酰 tRNA 一开始并非结合在 A 位点，而是结合在一个 T 位点 [图 8.26（a）]，氨酰 tRNA 与 EF-Tu*GTP 组成复合体的位点阻碍

了 tRNA 接近 A 位点。但是，对结合到小亚基上的反密码子臂的晶体学研究发现，反密码子恰当地结合在了 A 位点上。高分辨率冷冻电镜研究证实了在对密码子识别的过程中 tRNA 上生出一个“扭结”（kink），这个结位于反密码子臂和 D 臂之间 [图 8.26（b）]。

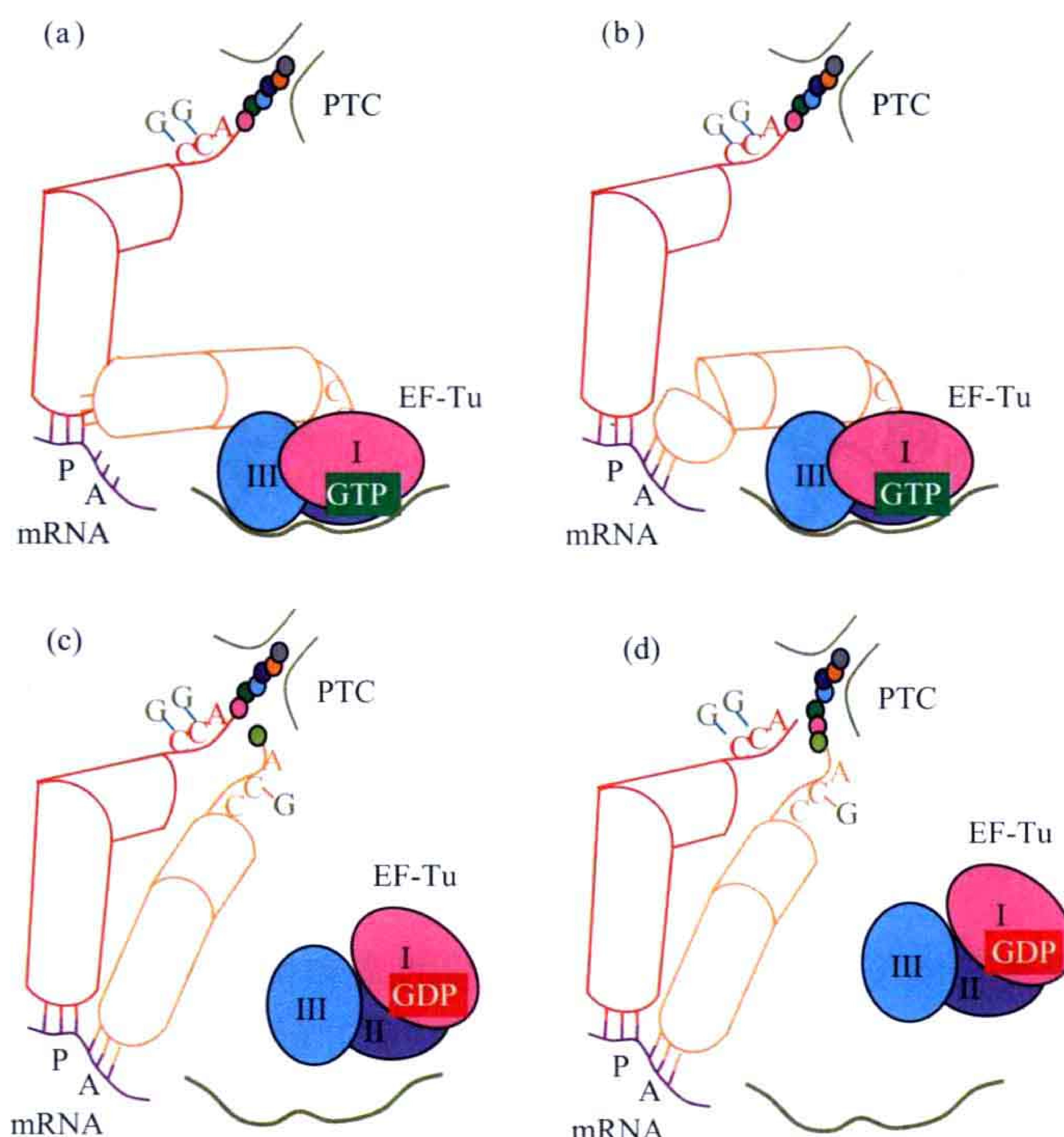


图 8.26 氨酰 tRNA 结合到 A 位点的主要步骤。(a) P 位点由一个通过末端 CCA 连接着一段肽段的 tRNA 占据。核糖体 RNA 上的两个 G 通过与 tRNA 末端两个 C 的配对将 tRNA 末端固定在这个位点上。氨酰 tRNA、EF-Tu 及 GTP 形成的三聚复合物结合在核糖体的 T 位点上。密码子尚未与反密码子相互作用。(b) 通过反密码子臂和 D 臂之间产生的一个扭结，反密码子得以与密码子相互作用。(c) 如果密码子与反密码子正确配对，EF-Tu 就会收到一个水解 GTP 的信号。之后 EF-Tu*GDP 复合体脱离 tRNA，tRNA 得以进入 A 位点，tRNA 上的氨酰残基也得以靠近肽酰转移中心（PTC）。23S RNA 的一个 G 可以与 tRNA 的 C75 相互作用形成碱基对。(d) 肽酰转移得以发生。

在这步起始选择中，核糖体也参与了鉴定正确配对和错误的相互作用。密码子 - 反密码子对上的第一和第二对碱基的 Watson-Crick 碱基配对的正确性是由位于小亚基的螺旋 h44 和 G530 上的两个腺嘌呤——1492 和 1493 完成的（图 8.27）。小亚基的螺旋 h44 和大亚基的螺旋 H69 之间的紧密接触能够调控水解 EF-Tu 上的 GTP 的信号，进一步导致 EF-Tu 脱离 tRNA 并使 tRNA 进入 A 位点 [图 8.26（c）]，但在 tRNA 进入 A 位点时其仍有掉离核糖体的危险，该机制被称为校对（proofreading）。

8.5.4 肽酰转移

肽酰转移的过程并不需要延伸因子的催化作用。当氨酰 tRNA 进入 A 位点后，

tRNA 携带的氨基酸被放置到肽酰转移中心（PTC），在 PTC 的旁边是伸向大亚基多肽出口通道的正在合成的多肽段 [图 8.26（d）]。PTC 主要由 23S RNA 组成，然而，在细菌中，核糖体是一个核酶的假设未必成立，因为在细菌核糖体中蛋白质 L27 的 N 端会接近位于 P 位点的合成中的多肽段和 tRNA 间的酯键（图 8.28）。

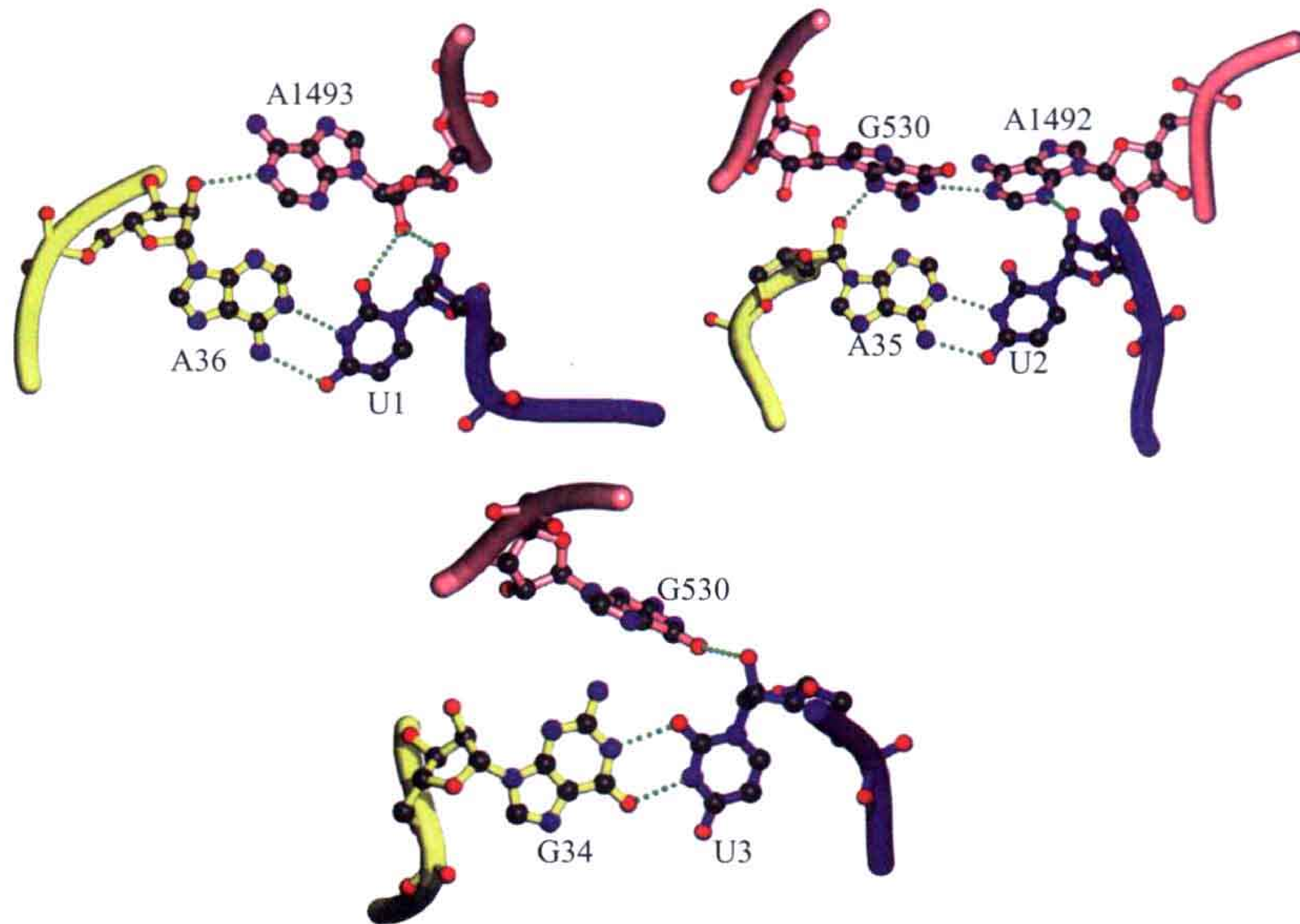


图 8.27 翻译的忠实性来自于密码子 - 反密码子间的正确相互作用。核糖体通过其小亚基 RNA 参与到这个过程中。mRNA 由蓝色表示；tRNA 的反密码子由黄色表示；核糖体 16S RNA 由粉色表示。前两位密码子 - 反密码子间的 Watson-Crick 碱基配对由 16S RNA 的 A1493、A1492 和 G530 进行检查（见左上图和右上图，PDB：2J00）。只有正确的密码子和反密码子间的配对才允许氢键的形成。对于非 Watson-Crick 配对，如 U-G 配对，A1493 无法与密码子核糖正确相互作用。在这种情况下 tRNA 不会与核糖体形成稳定的结构并会脱离下来。密码子第三位碱基的相互配对是摆动（wobble）的，并不是十分严格的配对（图中是一个 G-U 配对）（下图，PDB：1IBL）。

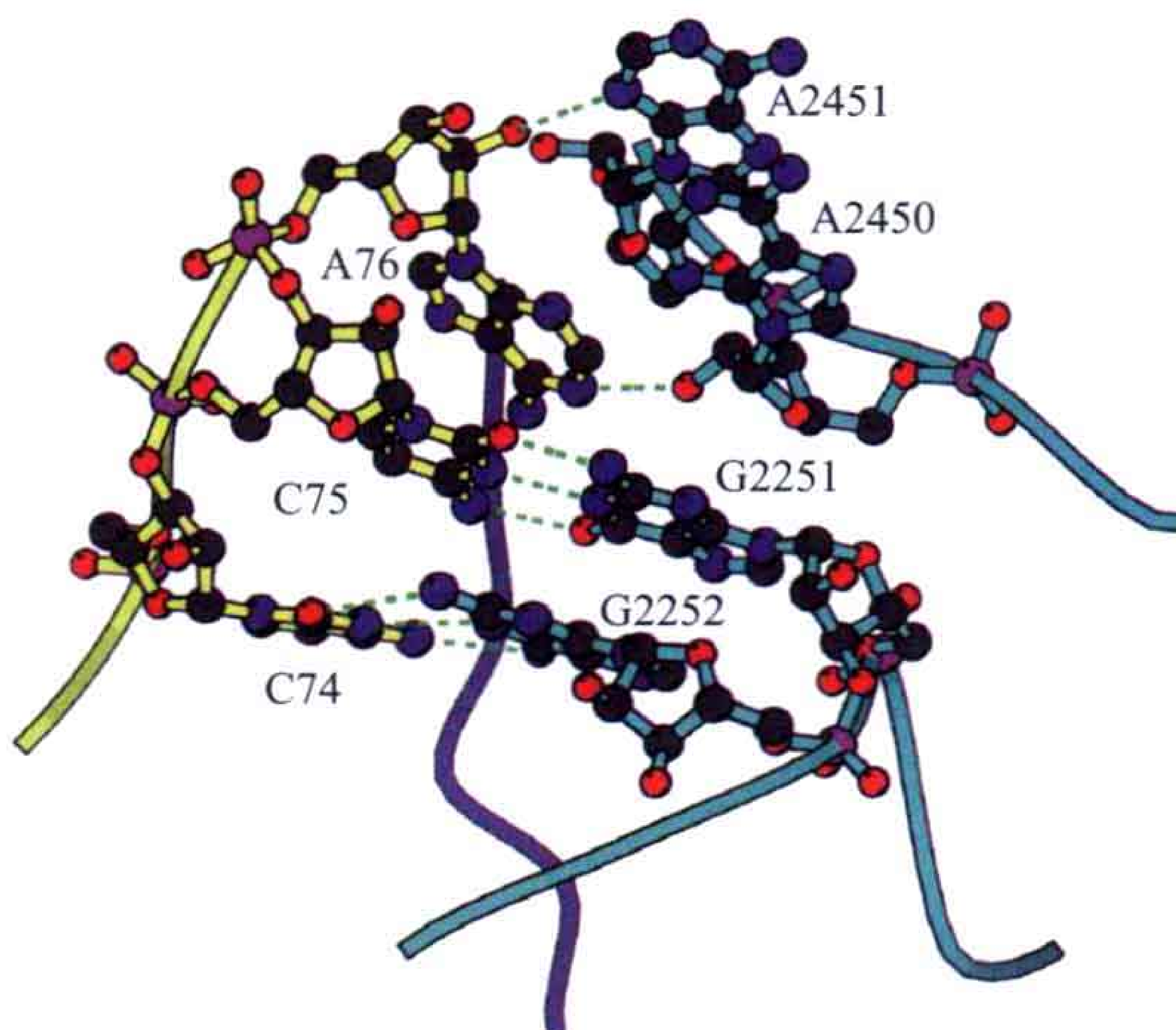


图 8.28 P 位点 tRNA 的接受末端（黄色）和 PTC 的 P-loop（绿色）。tRNA 3′ 端 CCA 中的两个胞嘧啶与 23S RNA 的两个鸟嘌呤形成碱基配对。末端的腺嘌呤与 3′ 端的其他碱基发生碱基堆积作用，并与 23S RNA 的另一个 loop 上的一个 2′ 羟基发生氢键相互作用。暗蓝色的一条链为蛋白质 L27 的 N 端，该蛋白质已知对有效的肽酰转移非常重要（PDB：2J00 和 J01）。

肽酰转移的主要步骤如图 8.29 所示。A 位点 tRNA 携带的氨酰残基的氨基基团分别与 P 位点 tRNA 的 A76 的 2′OH 间和 A2451 间形成氢键结合，这使得该氨基基团具有可以进攻 P 位点正在合成的肽段的羰基碳的走向和活性，从而使肽酰转移得以发生。

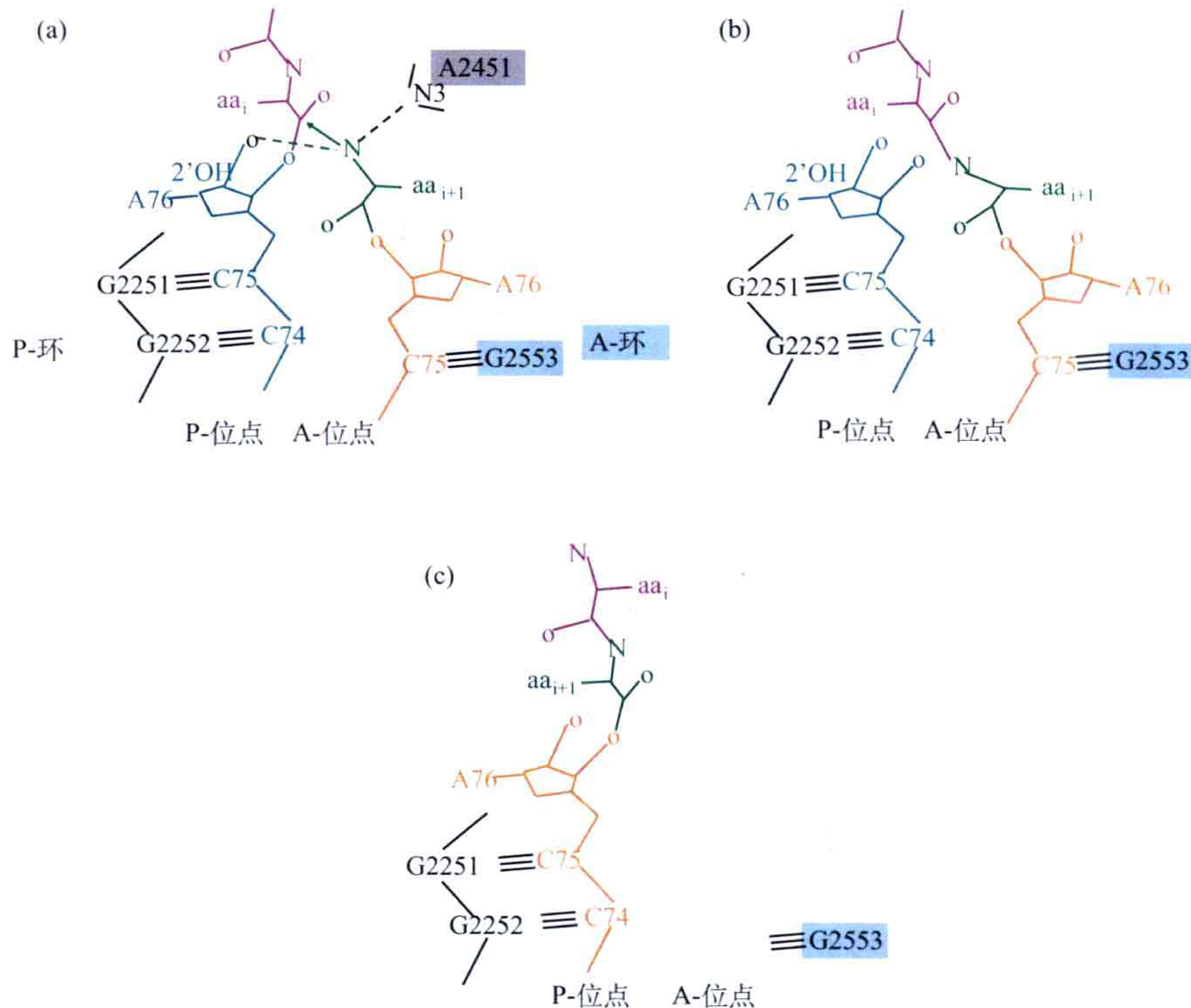


图 8.29　肽酰转移的主要步骤。(a) A 位点及 P 位点的 CCA 末端分别能与 23S RNA 的两条 loop 发生碱基配对。正在合成的多肽链定位在出口通道，A 位点 tRNA 携带的氨酰残基的氨基基团通过与 A2451 和 P 位点 tRNA 末端腺嘌呤的 2′OH 间形成两个氢键相互作用。(b) 肽酰转移发生。(c) 新合成的肽酰 tRNA 的 CCA 末端代替了结合在 P-loop 上的脱酰化的 tRNA，并同时结合到 A 位点和 P 位点，即一个 A/P 位点。

正在合成的多肽是定位在出口通道的，尽管肽酰转移后它连接到了 A 位点 tRNA 上，但本质上它不会跳到另一个位点而是一直待在同一位点。这使得 tRNA 结合在一个交叉位点（hybrid site）上，即 A/P 位点。这需要原来 P 位点的 tRNA 先进入一个 P/E 位点。引起的主要变化是来自于 A 位点 tRNA 的单链 CCA 末端的 180°重新定向。肽酰 tRNA 的 CCA 末端的 180°旋转得益于 PTC 的二次对称性。在 A 位点时 tRNA 的 C75 与 23S RNA 一个 loop 的 G-2251 碱基配对，而进入 P 位点后，它是与另一个对称 loop 的 G-2251 配对（图 8.28 和图 8.29）。在广泛研究到底哪些 A 位点核苷酸与 P 位点核苷酸有这种对称作用关系后，发现在每个位点大约有 90 个核苷酸具有这种关系。

8.5.5　移位

延伸循环的最后一步是移位（translocation），移位是由 EF-G 催化完成的。正如上

面讨论的，位于 A 位点肽酰 tRNA 的接受末端已经自发地旋转 180° 进入 P 位点，产生一个处于 A/P 位点的肽酰 tRNA，tRNA 的大部分依然留在 A 位点。EF-G*GTP 复合体模拟了 aa-tRNA*EF-Tu*GTP 三聚复合物的结构，并与之有相同的核糖体结合位点（图 8.30）。与核糖体结合了的 EF-G 的构象与没有核糖体结合的 EF-G 的晶体结构不同（图 8.31）。

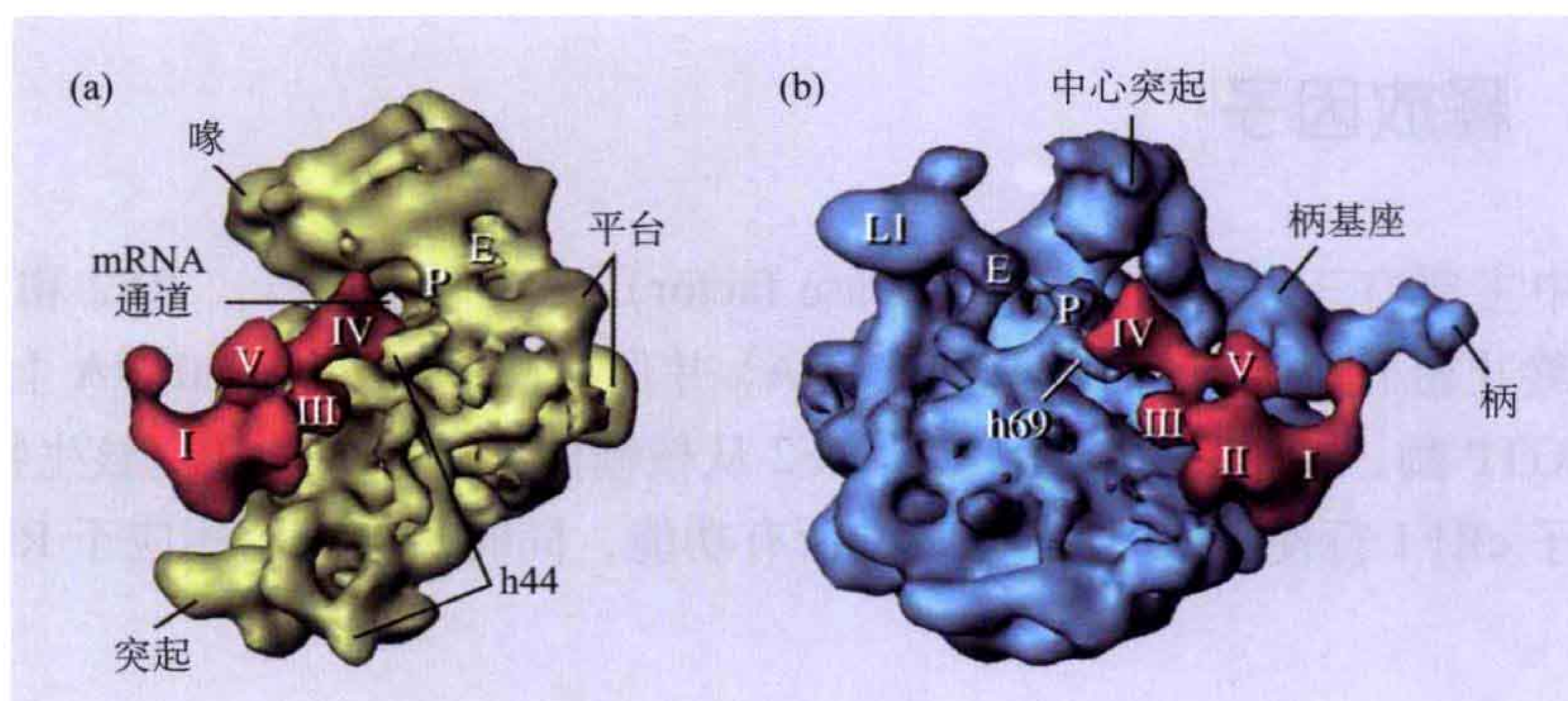


图 8.30　结合在核糖体上的 EF-G（红色）的冷冻电镜研究。结构域 I 和 II 与 EF-Tu 相应的结构域在大亚基上有相同的结合位点。具有类似反密码子结构的顶端部分（结构域IV）进入 A 位点并使肽酰 tRNA 从 A 位点移至 P 位点 [经 Frank J, Agrawal RK 授权再印制（2001）Ratchet-like movements between the two ribosomal subunits: Their implications in elongation factor recognition and tRNA translocation. Cold Spring Harbor Symp Quant Biol **LXVI:** 67-75. 版权 (2001) Cold Spring Harbor Laboratory Press.]。

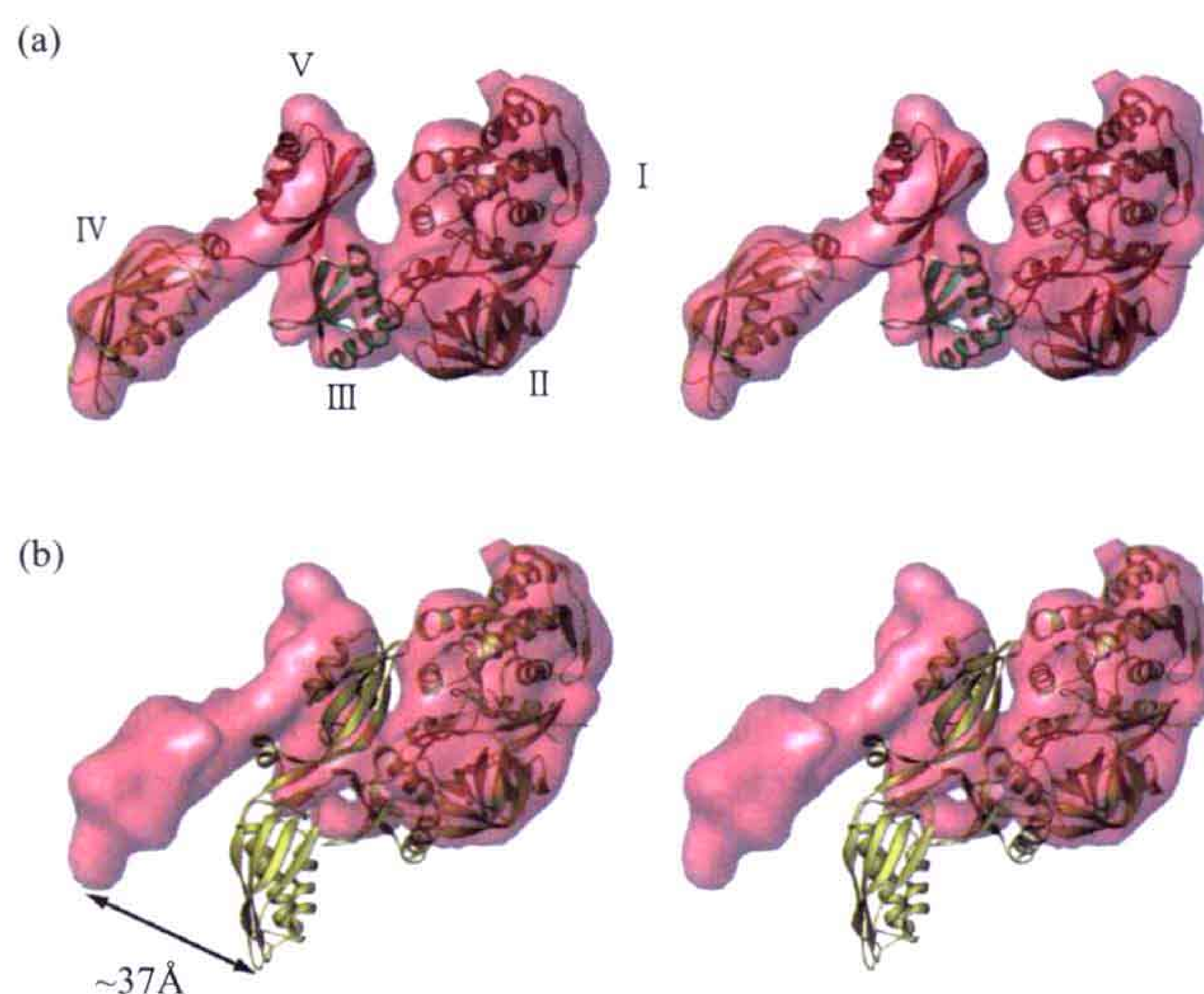


图 8.31　EF-G 的晶体结构与冷冻电镜密度图，并将 EF-G 晶体结构模型搭建至冷冻电镜密度图。反密码子臂（ASL）的顶端必须移动约 37Å 才能与结合核糖体时的冷冻电镜密度相匹配 [经 Valle M 等授权再印制（2003）Locking and unlocking of ribosomal motions. Cell **114:** 123-134. 版权（2003）Elsevier.]。

EF-G 需要与一个 GTP 分子结合才能发挥活性。但对 GTP 在移位之前还是移位之后被水解尚不清楚。如果在移位后被水解，EF-G 表现得更像一个经典 G 蛋白和分子开关；如果水解发生在移位前，EF-G 的行为则更像一个消耗 GTP 的马达蛋白。

8.6 多肽释放及核糖体再循环

8.6.1 释放因子

在细菌中主要有三种释放因子（release factor），它们是 RF1、 RF2 和 RF3。RF1 和 RF2 识别终止密码子（UAA、UAG 和 UGA）并促进肽段从 P 位点 tRNA 上水解下来。RF3 是一个 GTP 酶，并能够催化 RF1 和 RF2 从核糖体上的释放。在真核生物中一个单独的释放因子 eRF1 行使了 RF1 和 RF2 的所有功能，同时具有一个相应于 RF3 的 eRF3（表 8.5）。

表 8.5 原核生物和真核生物中的释放因子

细菌	终止密码子	真核生物	功能
RF1	UAA，UAG	eRF1	识别终止密码子并水解多肽段
RF2	UAA，UGA	eRF1	识别终止密码子并水解多肽段
RF3	—	eRF3	将其他的释放因子再循环

真核释放因子 eRF1 负责识别 mRNA 上的终止密码子，并能促进肽段从 tRNA 上水解。因为 mRNA 和合成的多肽段位于 P 位点 tRNA 上不同的两端，要同时与这两部分相互作用就要求 eRF1 至少与 tRNA 分子有相同的大小甚至形状。人 eRF1 是一个有着三个结构域的伸展结构（图 8.32）。无论在真核生物还是细菌中，在结构域 2 的末端都发现了一个共同的 GGQ 模体，该模体对核糖体 23S RNA 上的肽酰转移中心发挥的水解活性非常重要。eRF1 的另一个末端由两个螺旋结构构成，突变研究证明这一端能够结合 mRNA 上的终止密码子。大多数 RNA 结合蛋白都是基于反平行 β 折叠片的结构，因此 eRF1 具有一段特殊的结合 RNA 的构象。在一定程度上，eRF1 分子与 tRNA 有相似的大小和形状，它能够与 A 位点的终止密码子结合，并能与 P 位点带有肽段的 tRNA 相互作用。因此这是另一个具有 tRNA 类似结构的蛋白质的例子。

细菌释放因子 RF1 和 RF2 行使与 eRF1 类似的功能。它们同样含有参与水解和释放的相同的 GGQ 模体。但是 RF2 与真核释放因子没有序列或结构上的相似性。从晶体结构看，GGQ 模体和已知与终止密码子的解码相关的序列的距离不超过 23 Å，该距离比小亚基解码位点和大亚基 PTC 间的距离要短得多。但是，当结合到核糖体上时，该蛋白质的构象却表现出了极大的不同（图 8.32）。几个独立的结构解析和研究已经证实了当结合到核糖体上后该蛋白质存在一个大的构象变化，而对于这一点在功能上的重要性还需要进行进一步的研究。

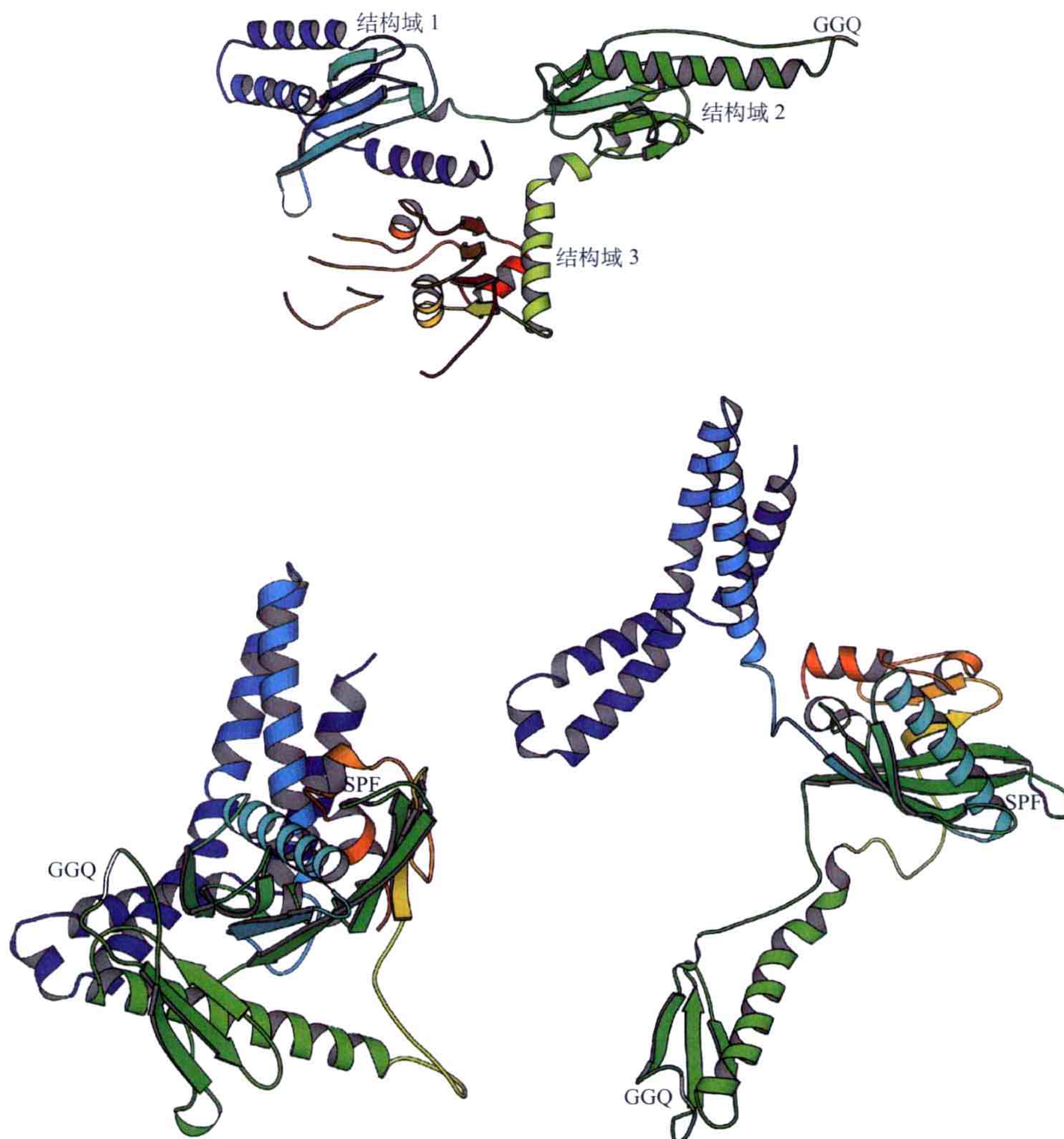


图 8.32　上：含有三个结构域的人 eRF1，与多肽水解相关的 GGQ 序列位于结构域 2 的顶端，反密码子可能由结构域 1 左侧的螺旋结构识别（PDB：1DT9）；下：细菌 RF2 的独立结构（左，PDB：1GQE）及位于核糖体上时的结构（右，PDB：1MI6）。GGQ 模体（白色）和位于解码位点的 SPF 模体（紫色）的距离在两种状态下截然不同。

8.6.2　核糖体再循环因子

在多肽段释放后，mRNA 和脱酰化的 tRNA 仍然结合在核糖体上，该复合体的解聚需要另一个蛋白质——核糖体再循环因子（ribosome recycling factor，RRF）的帮助。该蛋白质的构象极不寻常，它由两个结构域组成。结构域 I 由三条长条形且几乎平行的螺旋结构组成；另一个结构域由 N 端的螺旋结构之后的序列组成。整个分子呈现出 L 形结构，而且它的大小和形状与 tRNA 分子十分接近（图 8.33）。

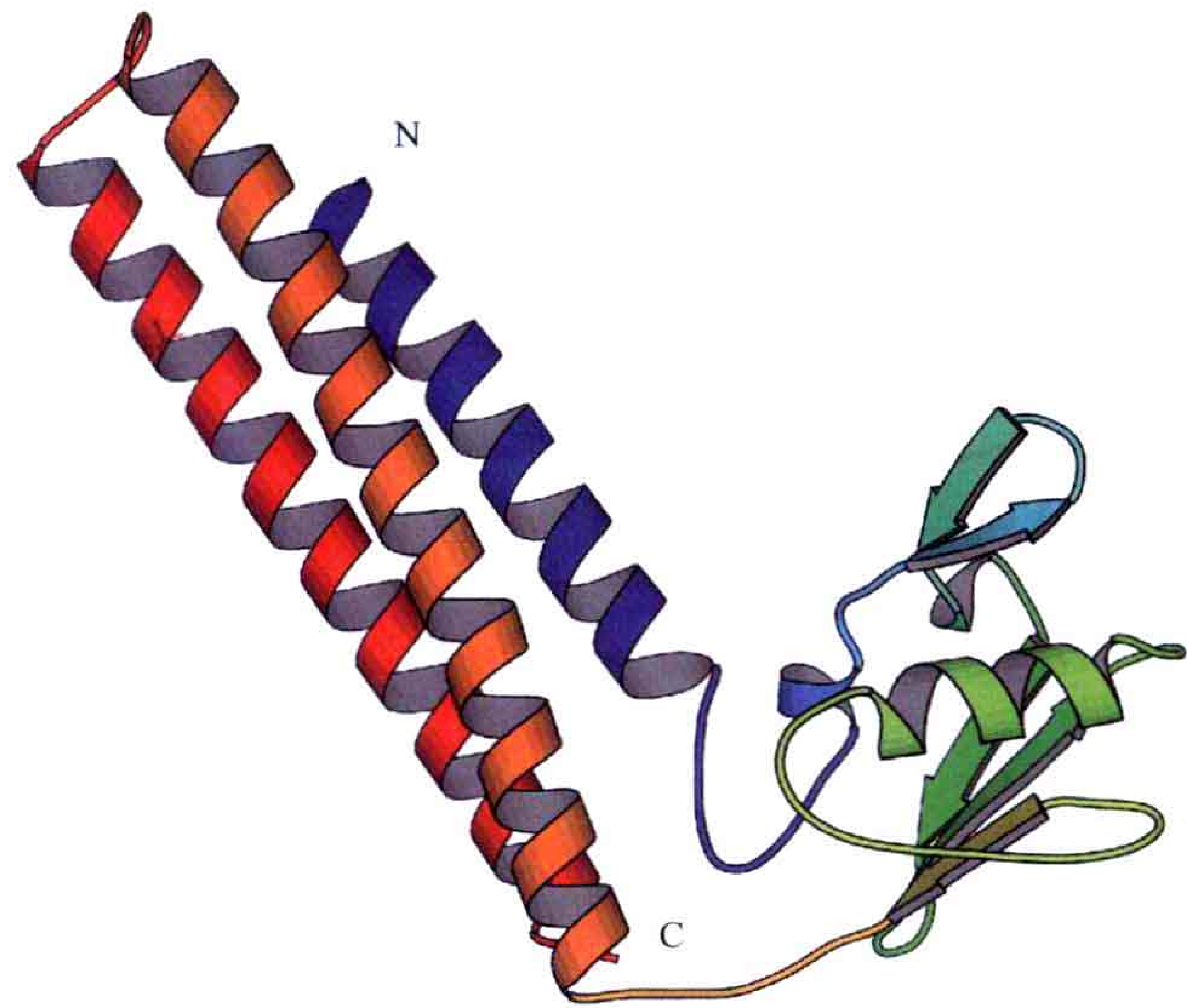


图 8.33　来自 *Thermotoga maritima* 的核糖体再循环因子（PDB：1DD5）。

多个 RRF 的结构已经得到了解析，这些结构证实了两个结构域间的转轴具有很强的柔性。RRF 是另一个具有类似 tRNA 结构的蛋白质，但是该蛋白质并不结合在核糖体上的一个单独的 tRNA，而是同时结合在 A 位点和 P 位点 tRNA 的接受臂部分（图 8.34）。

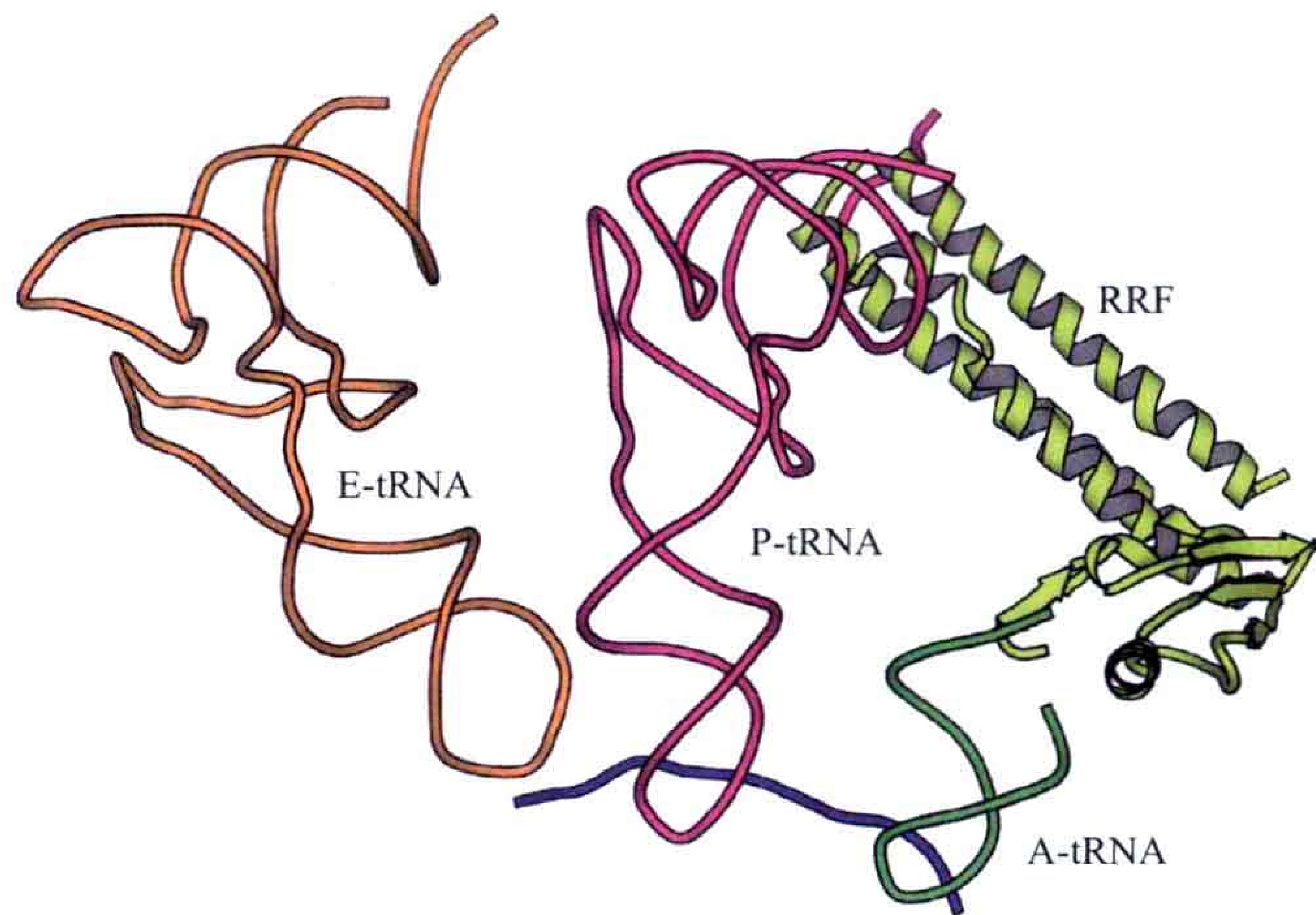


图 8.34　来自 *Thermus thermophilus* 的结合在 70S 核糖体上的 RRF 的晶体结构与结合 tRNA 的核糖体结构的重叠图(在这个结构中 A 位点的 tRNA 部分是无序的)。当 RRF 结合时没有 tRNA 结合在 A 位点。蛋白质的结合位点同时跨越了 A 位点和 P 位点。结合引发了一个构象变化，使 P 位点 tRNA 的接受臂移向 E 位点（PDB：2J00 和 2V46）。

RRF 需要 EF-G 和 GTP 的帮助才能行使核糖体解聚的功能。这可能需要 RRF 的移动或两个亚基的构象变化以打断核糖体亚基间的相互作用。再循环是翻译过程的第四步也是最后一步，之后 IF3 的结合阻止了亚基间不成熟的聚合。

延伸阅读 (8.2 节)

原始文献

Cavarelli J, Rees B, Ruff M, Thierry J-C, Moras D. (1993) Yeast tRNAAsp recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature* **362**: 181-184.

Nissen P, Thirup S, Kjeldgaard M, Nyborg J. (1999) The crystal structure of Cys-tRNACys-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. *Structure* **7**: 143-156.

Silvian LF, Wang J, Steitz TA. (1999) Insights into editing from an Ile-tRNA synthetase structure with tRNAIle and mupirocin. *Science* **285**: 1074-1077.

综述文章

Ibba M, Francklyn C, Cusack S (eds) (2005) *The Aminoacyl-tRNA Synthetases*. Landes Bioscience, Austin, Texas, USA.

延伸阅读 (8.3 ~ 8.6 节)

原始文献

Ævarsson A, Brazhnikov E, Garber M, Zheltsonova J, Chirgadze Y, Al-Karadaghi S, Svensson LA, Liljas A. (1994) Three-dimensional structure of the ribosomal translocase: Elongation factor G from *Thermus thermophilus*. *EMBO J* **13**: 3669-3677.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 resolution. *Science* **289**: 905-920.

Berchtold H, Reshetnikova L, Reiser COA, Schirmer NK, Sprinzl M, Hilgenfeld R. (1993) Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* **365**: 126-132.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. (2005) Structures of the bacterial ribosome at 3.5Å resolution. *Science* **310**: 827-834.

Selmer M, Al-Karadaghi S, Hirokawa G, Kaji A, Liljas A. (1999) Crystal structure of

Thermotoga maritima ribosome recycling factor: A tRNA mimic. *Science* **286**: 2349-2352.

Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**: 1935-1942.

Weixlbaumer A, Petry S, Dunham CM, Selmer M, Kelley AC, Ramakrishnan, V. (2007) Crystal structure of the ribosome recycling factor bound to the ribosome. *Nat Struct Mol Biol* **14**: 733-737.

Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. (2000) Structure of the 30S ribosomal subunit. *Nature* **407**: 327-339.

综述文章

Liljas A. (2004) *Structural Aspects of Protein Synthesis*. World Scientific, Singapore.

Moore PB, Steitz TA. (2003) The structural basis of large ribosomal subunit function. *Annu Rev Biochem* **72**: 813-850.

Noller HF. (2005) RNA structure: Reading the ribosome. *Science* **309**: 1508-1514.

Ogle JM, Carter AP, Ramakrishnan V. (2003) Insight into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci* **28**: 259-266.

（邢　栋 译　苏晓东 校）

第九章
蛋白质的折叠与降解

9.1 蛋白质的折叠

蛋白质合成是生物科学的主题。蛋白质只有正确折叠为特定的三维结构，才能行使其功能。数十年前，Anfinsen 的实验证实，蛋白质折叠是一个自发的过程，因为葡萄球菌核酸酶能够从变性的状态重折叠为有功能的状态。在折叠过程中，氨基酸序列决定蛋白质的折叠形式。通过分析氨基酸序列可以推断蛋白质的结构特征，也可能推断其折叠形式。但这仍然是一个没有被很好解决的基本问题。蛋白质似乎不太可能在折叠过程中起到帮助作用。怎么会有蛋白质能够为如此多的折叠形式当模子，而这些当模子的蛋白质又是怎样折叠的呢?

另外还有折叠的速率问题。Levinthal 指出，对于一个多肽，如果想搜寻所有可能的构象，需要的时间将长于宇宙的年龄。另外，细胞中总的蛋白质浓度约为 300 mg/mL，并且环境中存在多种蛋白酶，这对蛋白质是不利的。如果蛋白质没有正确的折叠，可能会聚集成团或被降解。然而，体外实验证实，单一结构域的蛋白质从变性状态到有功能的活性构象的折叠过程需要毫秒到秒的时间。显然，折叠过程一定通过某种途径来避免 Levinthal 猜想。多肽链上局部结构的形成可能起始了折叠过程。

9.1.1 自发的与受助的蛋白质折叠

自发的蛋白质折叠在第二章中已描述。很多蛋白质在其被合成出来的时候就自发折叠了。然而，一些蛋白质在细胞内条件下可能失去或者再也不能回复到它们原有的结构。因此在一些条件下，如升温，蛋白质需要得到帮助来获得它们有功能的结构。没有正确折叠的蛋白质可能聚集形成具有破坏性的淀粉样结构，这些淀粉样结构对生物体具有毒害作用。在用生物技术的方法过量表达蛋白质时，包涵体也常常会形成。然而，自然界已经发展出一系列补救机制来帮助那些需要帮助的蛋白质进行折叠。这些机制在整个演化过程中高度保守，暗示它们在生命系统中发挥至关重要的作用。

第一次观察到受助的折叠过程源于肽基脯氨酸顺反异构酶（PPIase）和蛋白质二硫键异构酶（PDI）的发现。某些蛋白质能特异地帮助脯氨酸残基处主链顺反构象的转换。当一些含有脯氨酸的蛋白质加入这些酶后，其折叠速率大大加快。

与之类似，细胞中发现了一些能够被热诱导的蛋白质，它们被称为热激蛋白（Hsp）。其中有一些能够挽救热变性的蛋白质。研究者们观察到这些热激蛋白大小各异，也发现不同物种之间的这些蛋白质具有同源性。研究者们猜测，这些蛋白质在细胞中可能不仅在胁迫环境下发挥功能，而且在正常条件下也具有作用。

9.1.1.1 脯氨酸顺反异构酶类

对于脯氨酸残基，不论顺式构象还是反式构象都是容许的，因为肽键扭角 ω 的值可以是 0° 或 180°（见 2.1 节）。一些脯氨酸需要采取顺式构象来使蛋白质的其他部分正确折叠，且脯氨酸的异构过程会大大降低蛋白质折叠速率。第一个脯氨酸顺反异构酶在 1984 年被分离出来。同年，免疫抑制剂 cycloporine A 的受体蛋白 cyclophilin 被发现。另一个免疫抑制剂 FK506 的受体也被分离出来。几年后，研究者们意识到这些受体与脯氨酸顺反异构酶类相同。脯氨酸顺反异构酶类与两种脯氨酸构象的晶体已经得到研究（图 9.1 和图 9.2）。

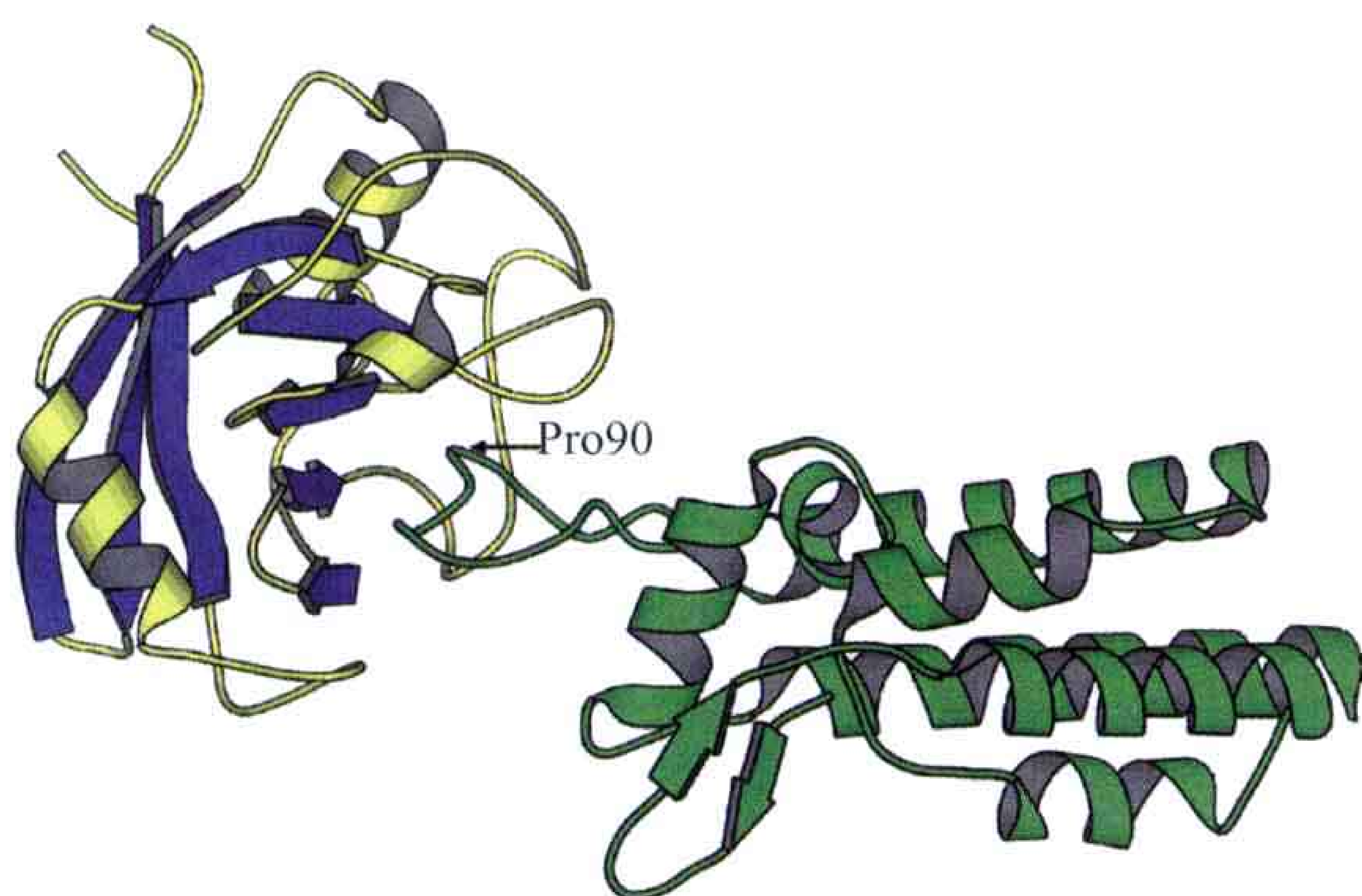


图 9.1 cyclophilin A（左）与 HIV-1 CA（衣壳）蛋白的 N 端结构域（右）的复合体结构。病毒蛋白中伸出的卷曲含有一个脯氨酸残基，这个残基与 cyclophilin A 相结合并经历了反 - 顺异构化的过程（PDB: 1M9C）。

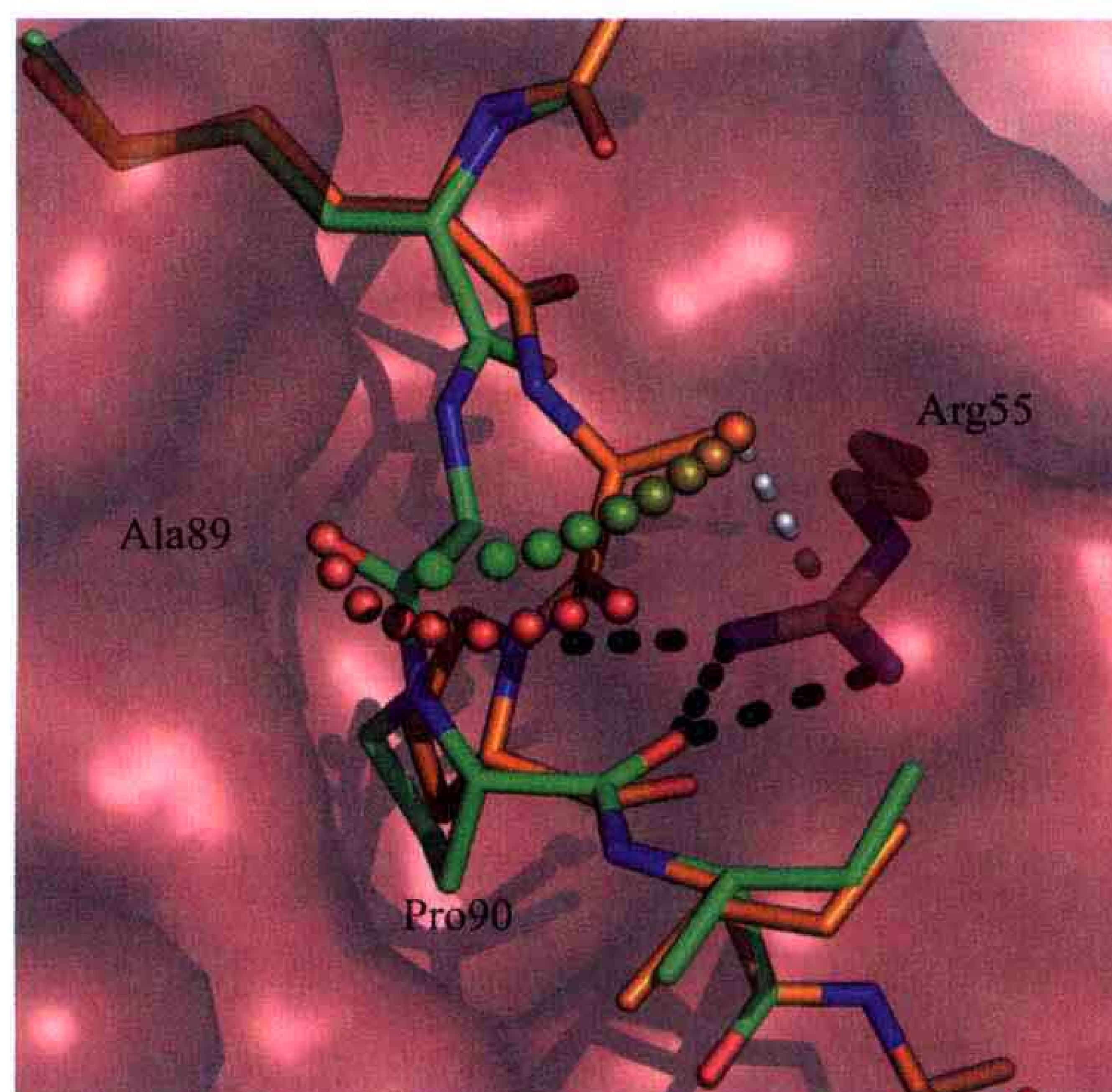


图 9.2 HIV-1 CA 蛋白第 90 位脯氨酸与 cyclophilin A 结合的精细图像。顺式构象和反式构象分别以绿色和橙色表示。第 89 位丙氨酸的 β 位碳原子和羧基氧原子的构象位移轨迹分别以绿色和红色小球表示。两个结构之间构象的差别主要限于第 89 位丙氨酸和第 90 位脯氨酸。cyclophilin A 的第 55 位精氨酸与 HIV-1 CA 蛋白第 90 位脯氨酸的羧基氧形成氢键，因此限制了第 90 位脯氨酸残基 C 端的构象变化。另外，精氨酸与成肽键的氮原子之间的氢键会稳定该脯氨酸氮原子金字塔形的 sp^3 杂化，也因此稳定了肽键的单键特征。这将有利于催化的进行 [重印经过许可，Howard et al., Structural insight into the catalytic mechanism of cyclophilin A. *Nat Struct Biol* : **10:** 475-481. Copyright (2003) Nature]。

除了羧基氧和脯氨酸前一个残基的侧链会发生很大的构象变化之外，由异构过程而导致的整体结构改变是很微小的。cyclophilin A 与 HIV-1 CA 蛋白之间的相互作用也因此而不同。催化过程是通过削减肽键的双键性质进行的。

9.1.1.2　蛋白质二硫键异构酶类

1963 年，Anfinsen 首次分离并表征了蛋白质二硫键异构酶类（PDI）。一个具有潜在二硫键的蛋白质需要经历硫原子的氧化过程来形成二硫桥。如果有多对半胱氨酸参与形成二硫键，它们之间可能形成错误的连接。另外，氧化状态的半胱氨酸会导致蛋白质聚集。为避免聚集或错误折叠，错误形成的二硫键需要被还原，并且需要尝试重新氧化蛋白质形成正确的折叠。在细菌与真核生物中，这个过程的实现有所不同。

在细菌中，DsbA 参与含有二硫键的蛋白质最初的氧化过程，而异构化是由 DsbC 完成的（图 9.3）。在真核生物中，氧化和异构化都是由 PDI 一个蛋白质完成的（图 9.4）。

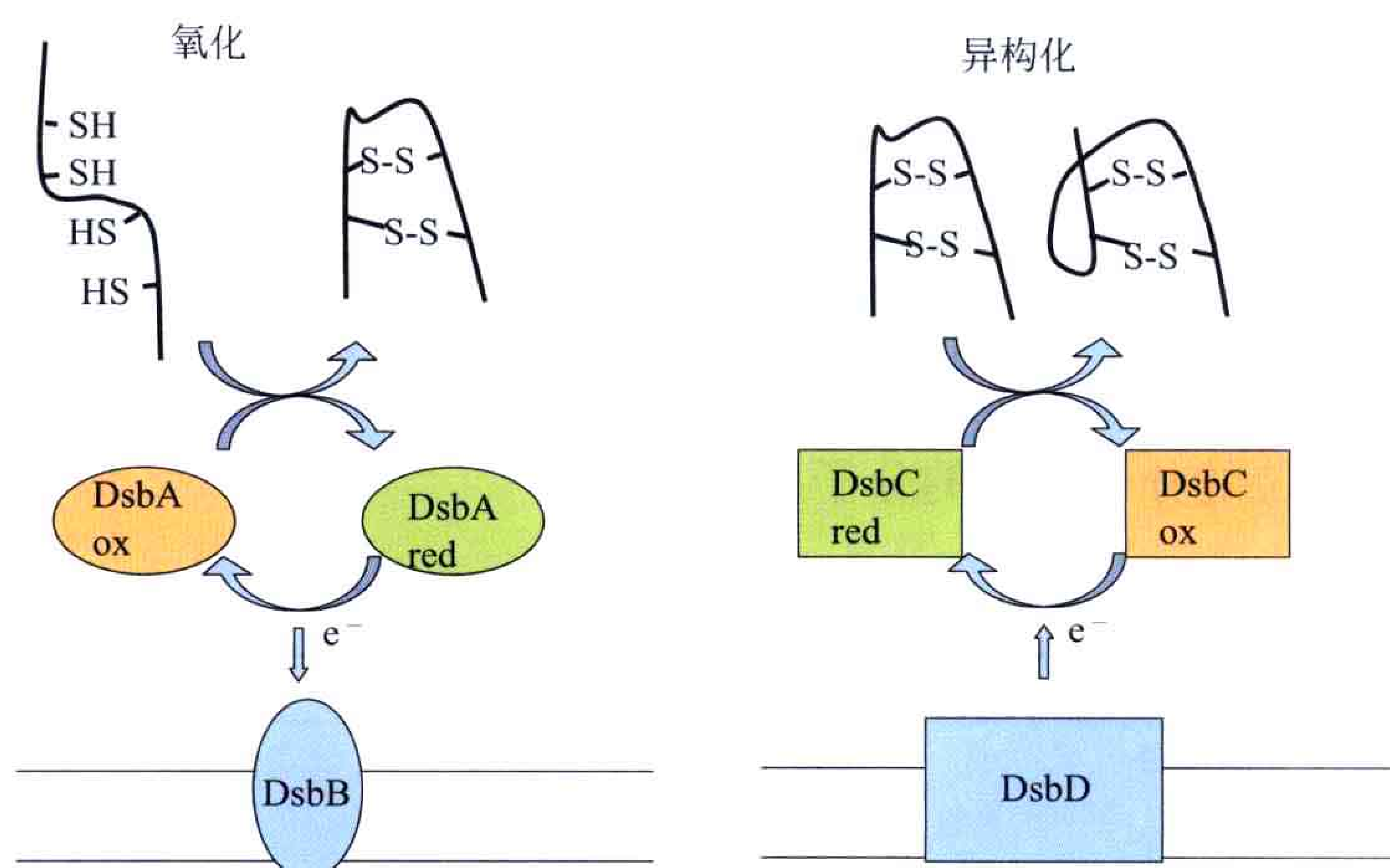


图 9.3　细菌中含有二硫键蛋白质的氧化和异构化过程。DsbA 蛋白负责氧化，DsbB 是参与重新氧化 DsbA 的一个膜蛋白。还原形式的 DsbC 催化异构化过程。DsbC 的还原态由 DsbD 维持。

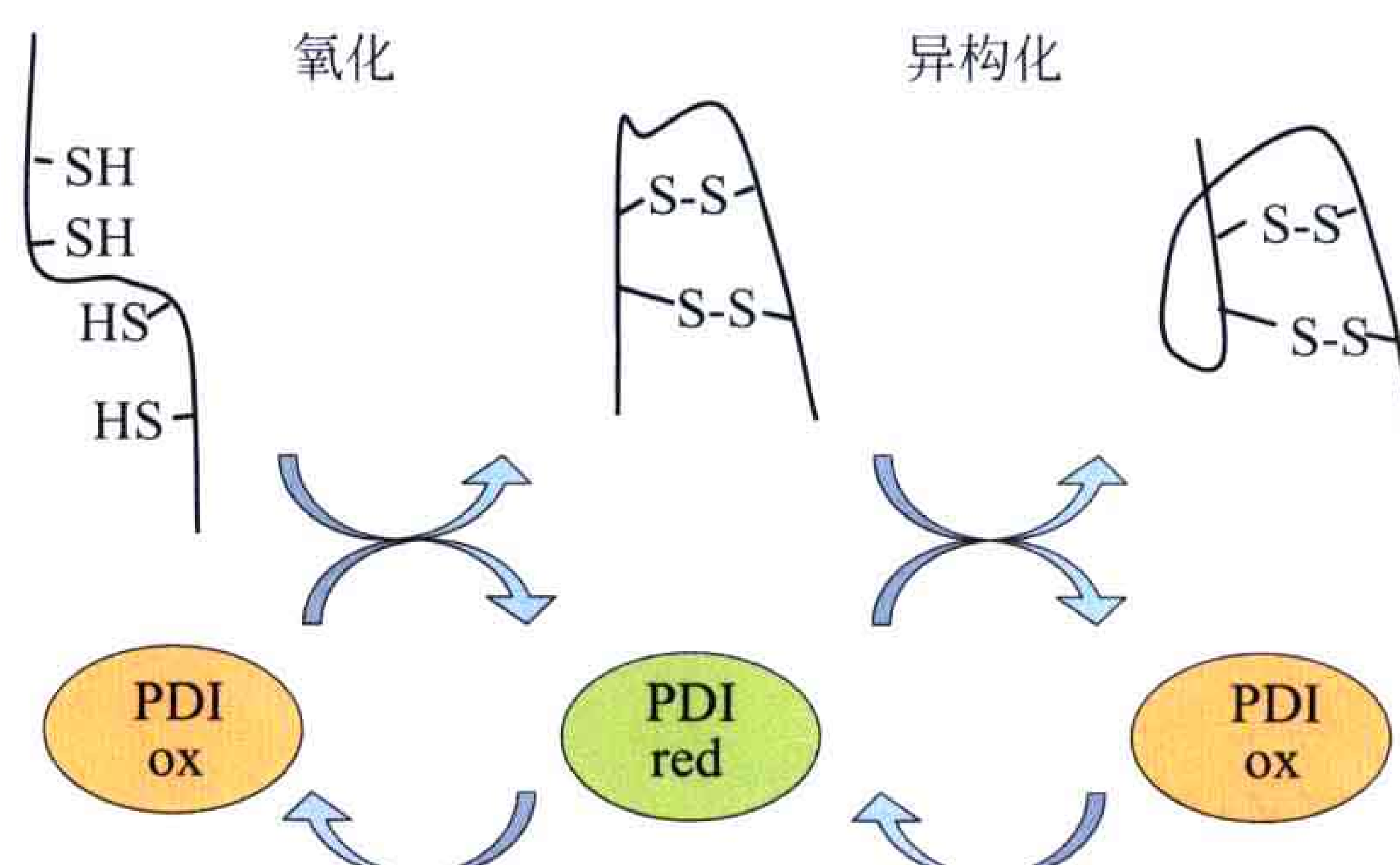


图 9.4　在真核生物中，PDI 同一个蛋白质在内质网中既负责氧化也负责异构化过程。PDI 的重新氧化由 FAD- 依赖的内质网的 oxidoreductin 蛋白完成。

行使氧化和异构化功能的蛋白质是由一个或数个硫氧还结构域组成的。CxxC 模体是结构域中参与催化的部分，这个部分中的两个半胱氨酸可以被还原或被氧化，形成一个二硫键。

9.1.2 分子伴侣

“分子伴侣”这个词于 1978 年第一次使用，它用于命名一个能够防止核小体错误装配的蛋白质。用这个命名的原因是，一个“伴侣”能够防止寡聚蛋白之间不合适的相互作用。后来研究者发现，数目不断增加的热激蛋白（heat shock protein, Hsp）类也具有正常的细胞学功能，因为在没有热胁迫时它们也存在于细胞中。这类蛋白质发现得越来越多并归类为分子伴侣。渐渐地，研究者了解到，分子伴侣不仅可以防止蛋白质有害的聚集，而且它们在许多蛋白质折叠的过程中也发挥重要作用，其中，一部分作用是使那些因受热或接触其他变性条件而变性的蛋白质复性。对分子伴侣的研究随后成为非常浩大且复杂的生物学研究的一部分。这一章将对分子伴侣一些基本的结构特性进行阐述。

热激蛋白类（表 9.1）主要分为三大类：小热激蛋白类（small heat shock protein, sHsp）、分子伴侣类（chaperon, 通常为单体蛋白）及伴侣素类（巨大的、中空的寡聚体结构）。后两大类蛋白质利用 ATP 来使部分变性的蛋白质复性。“Hsp”后面的数字表示以 kDa 为单位的分子伴侣的分子质量。

表 9.1 分子伴侣的一些代表性家族

名称	别称	功能	共同发挥作用的蛋白质
sHsp	小热激蛋白	结合非天然构象蛋白	Hsp40 + Hsp70，Hsp100
Hsp10	GroES	与 GroEL 共同形成 chaperonin	GroEL
Hsp40	DnaJ	结合疏水多肽	Hsp70
Hsp60	GroEL（伴侣素）	重折叠非天然构象蛋白	GroES
Hsp70	DnaK	重折叠非天然构象蛋白	Hsp40
		折叠新合成的蛋白质	
GropE	—	DnaK 的核苷酸交换因子	
Hsp78	线粒体 ClpB 同源物		
Hsp90	—	激活调控和信号蛋白	p23/Sba1
Hsp100	ClpA	使蛋白质聚集体重新可溶	ClpS
Hsp104	ClpB		ClpS
Hsp110	—	与 Hsp70 部分相关联	
TF	trigger factor	帮助新生多肽链折叠	核糖体

9.1.2.1 小热激蛋白类

在所有生物中存在着一大类类型多样的分子伴侣，称为小热激蛋白类（sHsp），它们的分子质量为 12~43 kDa。sHsp 能结合非天然构象的蛋白质并阻止它们聚集。

sHsp 常常在如热胁迫之类的胁迫条件下被诱导为活性状态。但这种激活的细节仍不清楚。sHsp 还参与许多基本的细胞活动，在临床上受到广泛关注。它们的突变与人类疾病息息相关，这些疾病包括白内障、肌肉疾病和神经疾病。例如，阿尔茨海默病和帕金森病等退行性疾病的细胞中频繁发生蛋白质聚集，而 sHsp 常在这种频繁发生蛋白质聚集的细胞中积累。

sHsp 的普遍特征是具有一个约 90 个氨基酸的保守的 α- 晶状体蛋白结构域（α-crystallin domain, ACD）。α- 晶状体蛋白（α-crystallin）是眼球晶状体的主要组成成分，它具有非常高的蛋白质浓度。αA- 和 αB- 晶状体蛋白在晶状体中很可能会阻止蛋白质聚集，而晶状体中的蛋白质聚集将导致视力障碍。有时一个 sHsp 单体内存在两个 ACD。ACD 的 N 端和 C 端外延，外延区域的长度和序列高度可变。α- 晶状体蛋白有一个类似于免疫球蛋白的折叠，是具有一个 β 片层的三明治结构。两个 β 片层分别有 3 和 4 个 β 束。α- 晶状体蛋白的最基本的排布是二聚体形式，其中，一个结构域的一个长的卷曲与另一个结构域的片层边界相互作用（图 9.5）。而二聚体之间的相互作用可以存在差异（图 9.6 和图 9.7）。

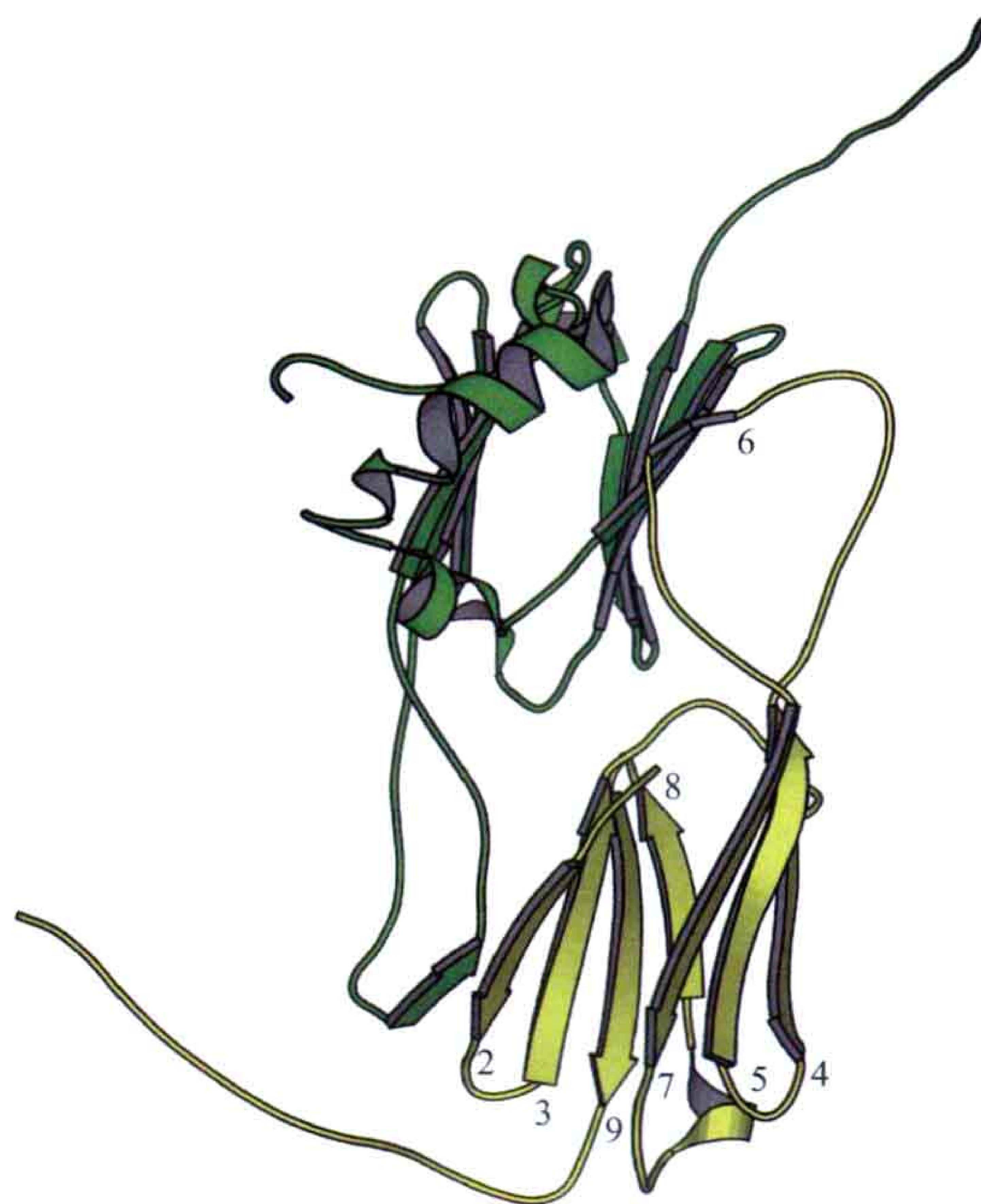


图 9.5　来自小麦的 16.9B 热激蛋白中，α- 晶状体蛋白结构域的组织形式。该结构域是一个 β 三明治结构，其中一个片层有 3 个 β 束而另一个片层有 5 个 β 束。结构的组织形式与免疫球蛋白折叠相关，但没有序列相关性。二聚体结构是以一种域交换形式存在的，其中一个单体的 β 束 6 使另一个单体的 β 片层得以延伸（PDB: 1GME）。

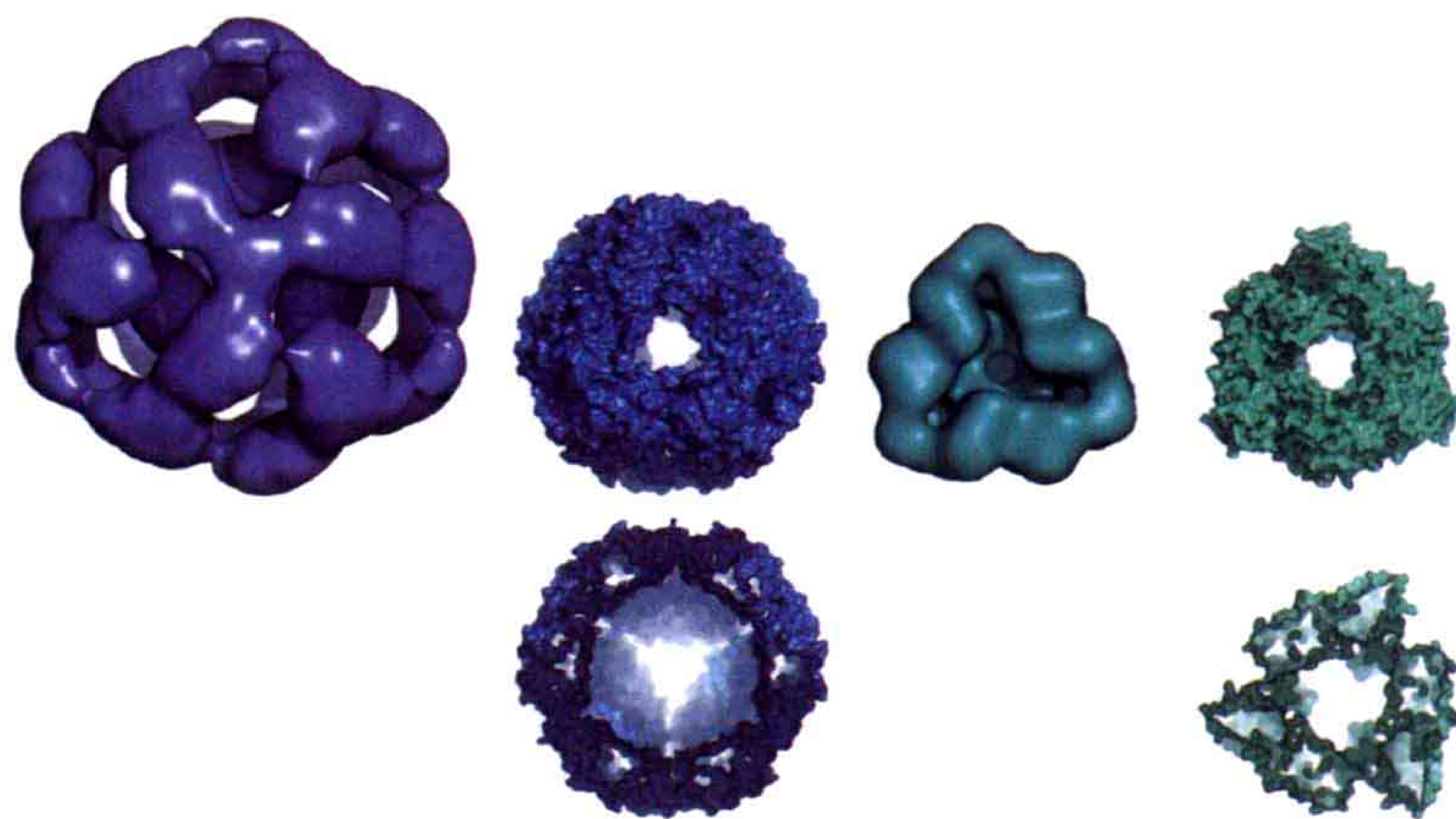


图 9.6　来自不同物种的 sHsp 二聚体可聚集为不同尺寸和不同亚基数目的中空实体。上：从左到右分别是 Hsp26、Hsp16.5、Acr1 和 Hsp16.9；下：显示的是 Hsp16.5 和 Hsp16.9 的截面图。图片由伦敦 Birkbeck 学院的 Helen Saibil 友情提供。

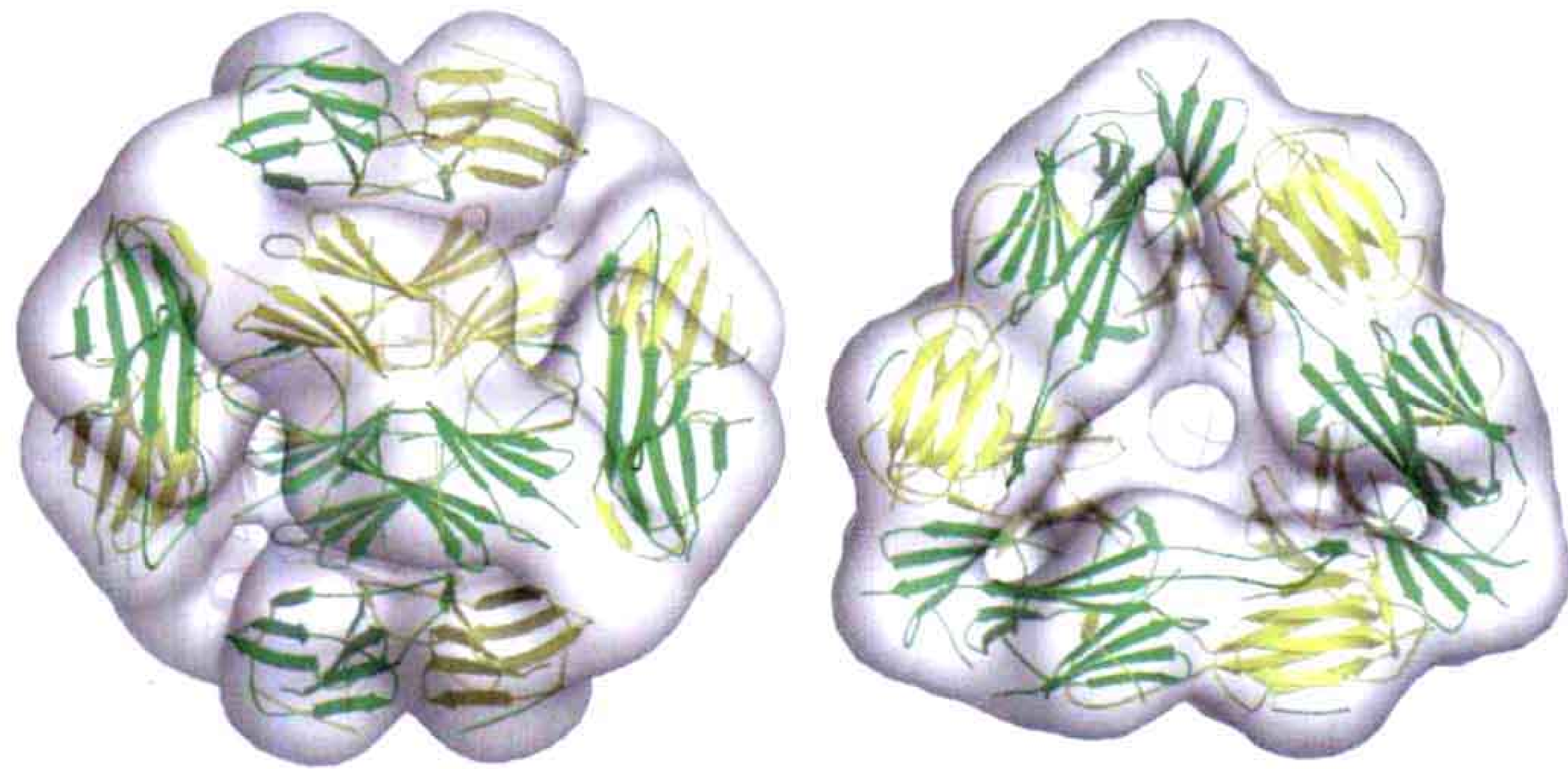


图 9.7 来自结核分枝杆菌（*Mycobacterium tuberculosis*）的 Hsp16.3（Acr1）十二聚体结构。α - 晶状体蛋白结构已填充于电镜负染所得的电子密度图中。左：沿二重对称轴向下观察；右：沿三重对称轴向下观察。图片由伦敦 Birkbeck 学院的 Helen Saibil 友情提供。

这些晶状体蛋白常常形成巨大的、中空的、动态的寡聚体。寡聚体的亚基数目为 12~50 个。蛋白质的末端区域似乎是柔性的，参与稳定寡聚体，且对分子伴侣的活性是必需的。sHsp 能和非天然构象的蛋白质形成大的复合体。它们也可以成为包涵体的组成部分，包涵体即一种巨大的胞内沉淀，当大量蛋白质在表达系统中被合成时，有时会形成包涵体。

sHsp 紧密结合非天然构象的蛋白质以阻止它们聚集或被酶解，尤其是在胁迫条件下。许多情况下，sHsp 聚集体在胁迫条件下通过构象改变而被激活，其底物的释放一般需要帮助。这些帮助是由一些更大的单体分子伴侣提供的，它们利用 ATP 来进行重折叠反应（图 9.8）。

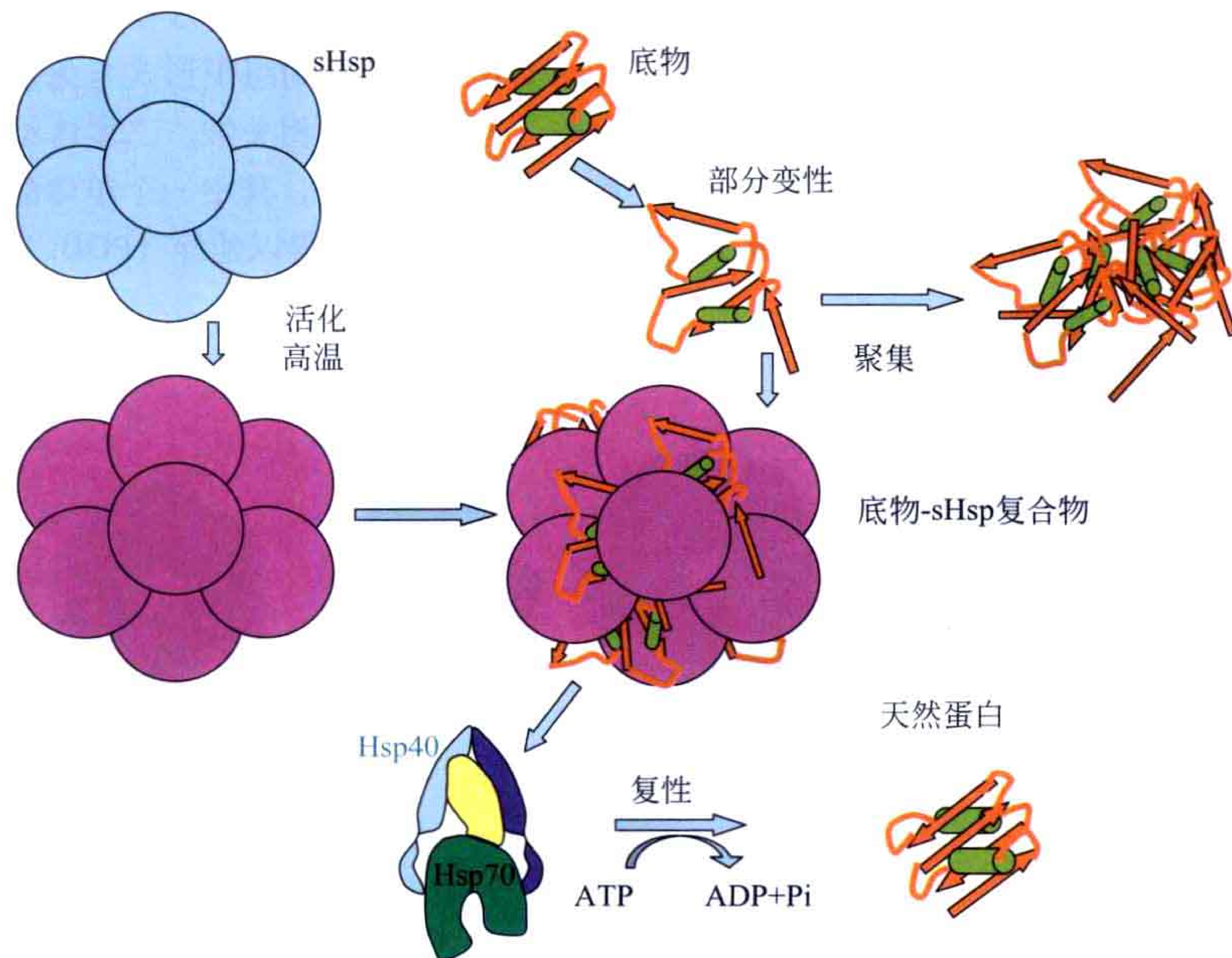


图 9.8 sHsp 常常以二聚体为单位装配成更大的寡聚体。这些复合体在胁迫条件下被激活并结合非天然构象的蛋白质。这种结合能够阻止非天然构象的蛋白质形成不可逆的沉淀聚集或被酶切降解。sHsp 寡聚体和底物蛋白形成的复合体能在如 Hsp40、Hsp70 和 Hsp100 单体分子伴侣的帮助下解聚。同时，Hsps 也能帮助底物分子进行重折叠。

9.1.2.2　分子伴侣类

许多分子伴侣以单体或二聚体形式存在。分子伴侣家族的 Hsp70 和 Hsp110 是单体，而 Hsp40 和 Hsp90 是二聚体。另外，Hsp100 属于 AAA+（ATPase associated with various cellular activity）超家族蛋白，它们是六聚体。分子伴侣结合和释放部分折叠的蛋白质依赖 ATP 的结合和水解。

9.1.2.3　Hsp70 和 Hsp40

许多分子伴侣需要与辅分子伴侣（co-chaperon）协作。Hsp40（在大肠杆菌中称为 DnaJ）是 Hsp70（在大肠杆菌中称为 DnaK）的辅分子伴侣。这类蛋白质在细菌和真核生物中广泛分布。Hsp40 以二聚体形式发挥功能，在两个单体之间有一巨大的裂隙。Hsp40 能结合未折叠的底物蛋白并阻止它发生聚集。Hsp70 随后与 Hsp40 的两个单体中的裂隙部分相互作用，并帮助未折叠的蛋白质折叠。Hsp40 有两种类型，分别称为 1 和 2。在两种类型的诸多不同点中，其中一点是类型 1 有两个锌指结构。Hsp40 的 N 端称为 J-结构域，它能激活 Hsp70 的 ATP 酶活性并调节 Hsp70 与底物的结合。这个结构域没有显示在所示的晶体结构中。Hsp40 的 C 端区域由两个或三个结构域组成。底物肽段以反平行的 β 束形式与结构域 I 中的 β 片层结合（图 9.9）。

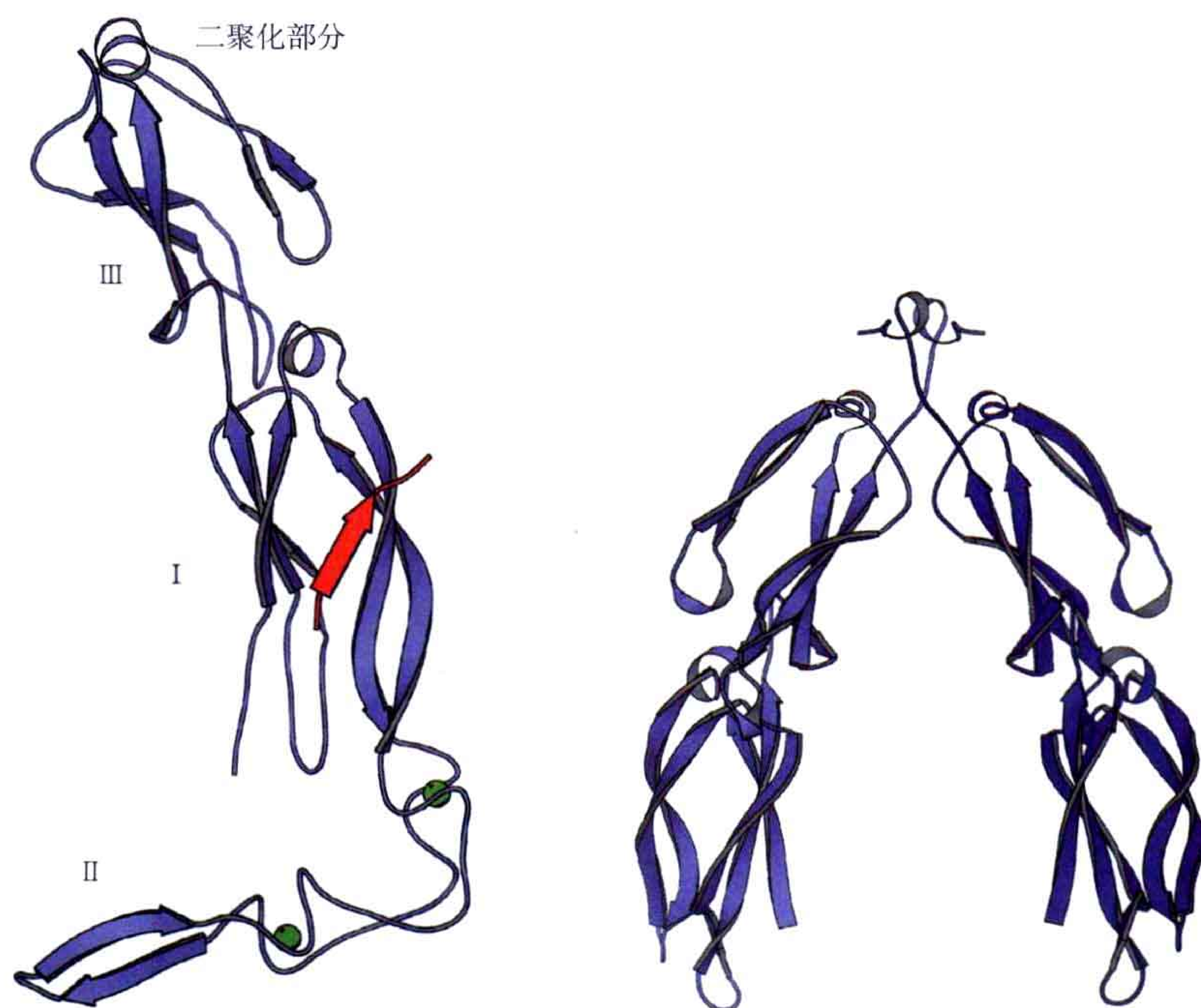


图 9.9　左：类型 1 的 Hsp40 结合有多肽的 C 端片段单体结构（来自酵母的 Ydj1，PDB:1NLT）。三个结构域 I、II 和 III 已标明。两个锌指结构和锌离子（绿色）是 Hsp40 类型 1 的特征。结合多肽的位置用红色表示。右：类型 2 的二聚体结构（来自酵母的 Sis1，PDB:1C3G）。这个蛋白质没有结构域 II。

Hsp70 是复合物中结合并水解 ATP 的组分。Hsp70 也与被称为 GrpE 的核苷酸交换蛋白进行相互作用。Hsp70 由一个 N 端的核苷酸（ATP）结合结构域和一个 C 端的底物结合结构域组成（图 9.10）。底物蛋白或多肽的结合与释放受 ATP 的结合与水解的调控，但细节尚不清楚（图 9.11）。

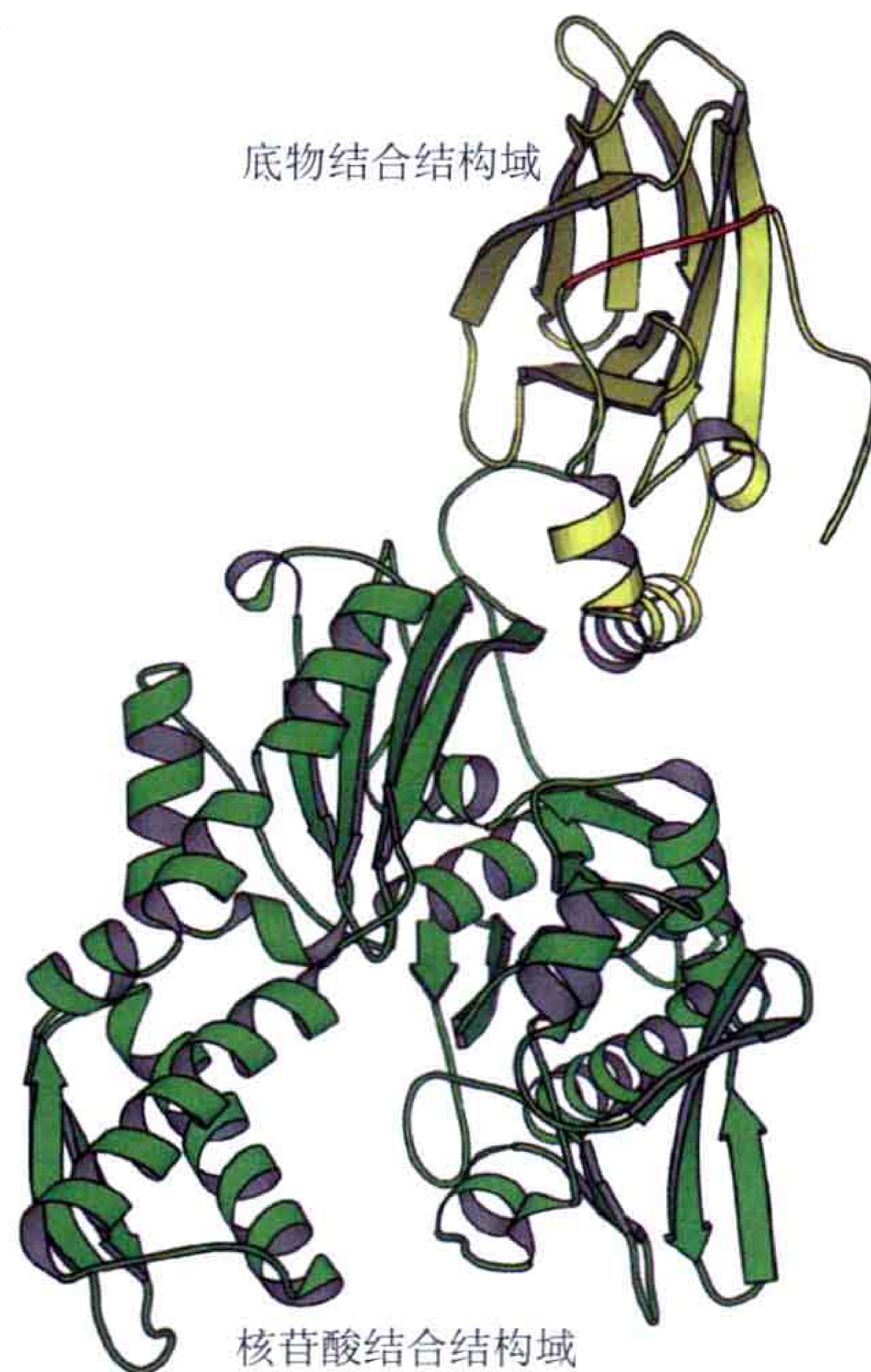


图 9.10　牛 Hsp70 的晶体结构，它属于分子伴侣类的 Hsp70 家族。核苷酸结合结构域以绿色表示。ATP 结合于分子底部的裂隙中。底物结合结构域以黄色表示。一段红色的多肽是 C 端的一部分，但是占据了底物结合位点（PDB:1YUW）。

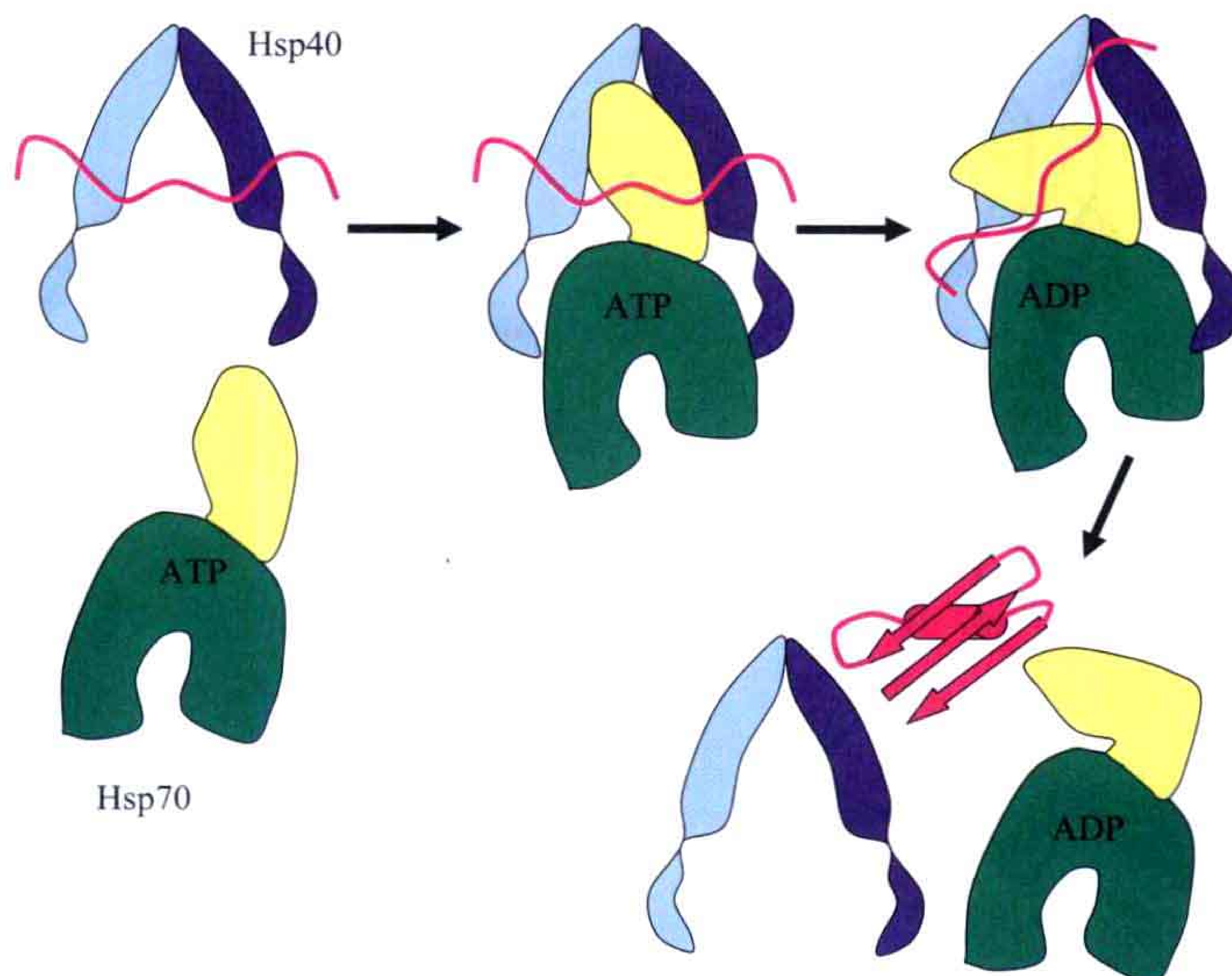


图 9.11　Hsp40 与 Hsp70 相互作用示意图。底物（红色）横跨结合于 Hsp40 二聚体上。当 Hsp40 的 J 结构域（目前尚无结构）与 Hsp70 相互作用时，ATP 被水解，并导致底物由 Hsp40 释放到 Hsp70 上。分子间的相互作用和构象改变只得到了部分了解。

9.1.2.4　Hsp90

Hsp90 是真核生物中二聚化的分子伴侣，它激活许多调控蛋白和信号蛋白，如原癌蛋白激酶。因此，它是抗癌药物研发的重要靶点。

长条状的单体由 4 个结构域组成（图 9.12）。位于末端的 N 端和 C 端结构域由 M 大结构域和 M 小结构域连接。ATP 结合位点在 N 端结构域。Hsp90 有大量辅分子伴侣。

来自酵母的 Hsp90 二聚体与两个辅分子伴侣 p23/Sba1 的结构已经被得到（图 9.13）。两个单体以二重轴为对称，参与底物或“顾客”蛋白相互作用的两性卷曲，从其中一个单体的 M 大结构域伸入 Hsp90 单体之间的空腔中，指向另一个单体的对应部分。单体间主要通过 C 结构域相互作用。N 结构域的相互作用有赖于 ATP 或其类似物的结合，结合通过结构域之间一个折叠片的交换进行。4 个 M 结构域之间形成一个空腔。这是“顾客”蛋白的结合区域（图 9.14）。

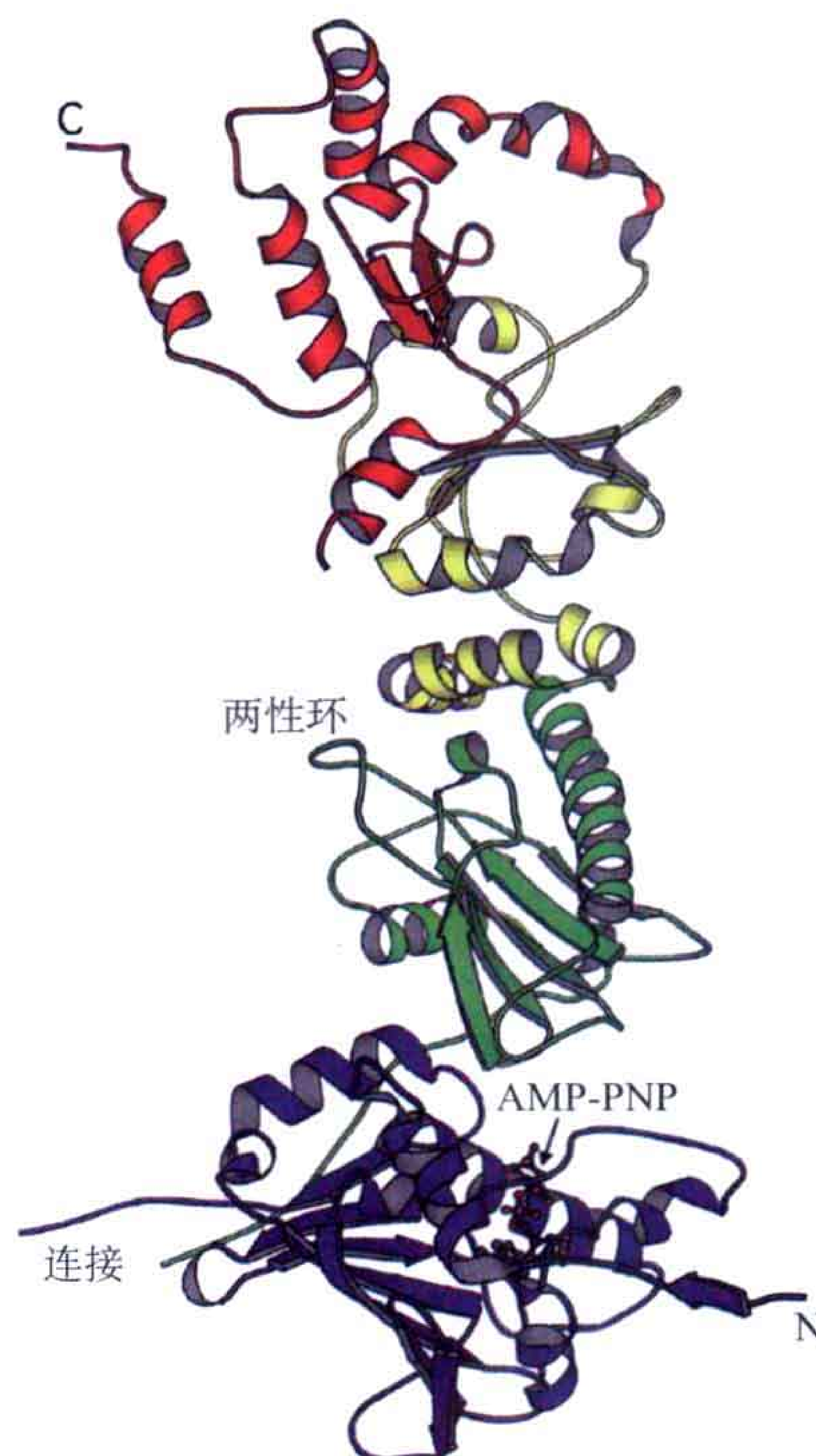


图 9.12　来自酵母的 Hsp90 结构。长条状的分子伴侣由 4 个结构域组成。ATP 结合结构域是 N 端的结构域（蓝色），然后是 M 大结构域（绿色）和 M 小结构域（黄色），最后是 C 结构域（红色）。N 结构域和 M 大结构域之间是通过一段突出的带电荷的序列连接的。两性的卷曲序列是“顾客”蛋白与分子伴侣相互作用的位置（PDB: 2CG9）。

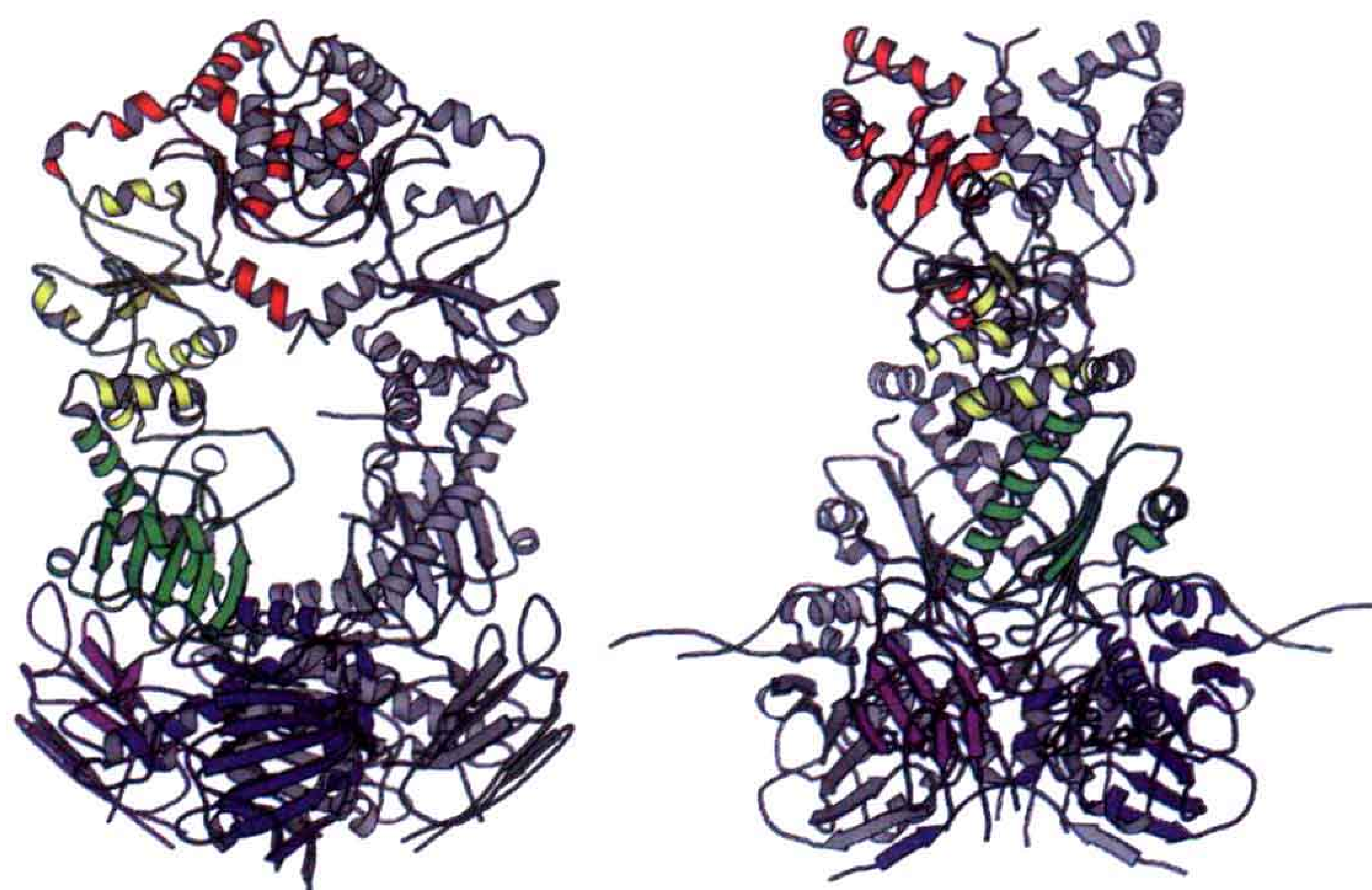


图 9.13　Hsp90 二聚体与其辅分子伴侣 p23/Sha1（紫色）的结构以两个互相垂直的视角显示。Hsp90 与 ADPNP 复合体的二聚化依靠 N 和 C 结构域的相互作用形成。N 端通过一个域交换来稳定结构。每个单体中有一个两性的卷曲，定位于 4 个 M 结构域之间的空腔中。辅分子伴侣结合于两个 Hsp90 的 N 端结构域之间。ATP 的结合使 N 端靠拢，并导致两个 M 大结构域产生 20Å 的移位。

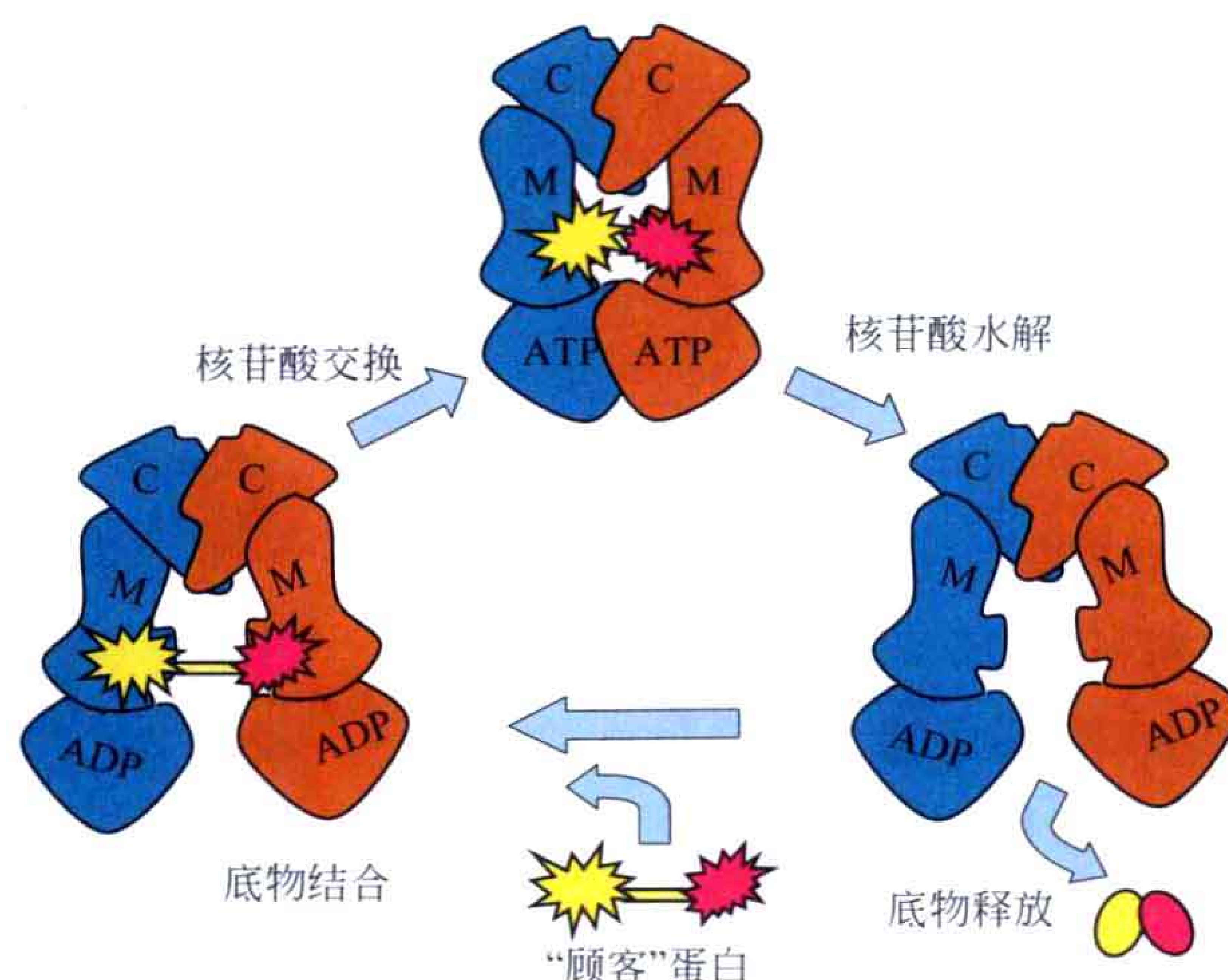


图 9.14　Hsp90 的一个功能模型。在与 ADP 结合的 Hsp90 中，N 端相互分离。ADP 被交换为 ATP 而引起了一个构象变化，从而使 N 结构域相互作用。在这个过程中，“顾客”蛋白被激活。随着 ATP 的水解，N 结构域之间的相互作用消失，激活的底物蛋白被释放。

ATP 或其类似物结合于 Hsp90 的 N 结构域会诱导构象变化。一个被称为盖子的结构在核苷酸附近闭合。这导致了一个疏水面的暴露，进而导致了两个 N 结构域的二聚化及与 p23 互作界面的产生。当 ATP 或 ATP 类似物结合于蛋白质上，并使蛋白质稳定在一种易于结合“顾客”蛋白的活性状态时，p23 优先与 N 结构域结合。底物或“顾客”蛋白能与开放的、ADP 结合构象的蛋白质结合。当 ADP 被 ATP 交换下来后，N 结构域将二聚化，并能结合辅分子伴侣。这将关闭 M 结构域之间的空间，并可能帮助底物蛋白构象产生变化。随着 ATP 的水解，盖子打开，N 端二聚化消失。二聚体结构的开放释放了有活性的“顾客”蛋白。

9.1.2.5　Hsp100

Hsp100 或 Clp 蛋白属于 AAA+ 超家族。这类 ATP 酶超家族有一个 P-loop 和一个类似于三层折叠的 G 蛋白，但是中心片层的拓扑结构并不相同。它们中的大部分是寡聚体并形成环状结构。

Hsp100 分子伴侣家族成员在细菌和所有真核生物中都存在。它们不仅可以作为分子伴侣去除蛋白质的聚集，也可以联合起来参与蛋白质的酶解。Hsp100 是同源六聚体。许多相关蛋白（Hsp104、ClpA、ClpB、Hsp78 等）都具有相同的特性：它们有两个 AAA+ 结构域，即 AAA-1 和 AAA-2（图 9.15）。

N 端结构域松散地连接在六聚体主体上，在如 ClpS 的小接头蛋白介导下参与底物的识别。ClpS 似乎可以引导 Hsp100 对聚集了的蛋白质发挥活性（图 9.16）。N 端结构域由两个螺旋束组成，每束有 4 个螺旋，两个螺旋束间存在一个假二重轴。N 端结构域存在一个疏水面。ClpS 与 N 端结构域通过多个接触位置相互作用。N 结构域的黏性面对底物的非特异性结合可能有贡献，但功能性结合在 AAA-1 和 AAA-2 结构域内发生。

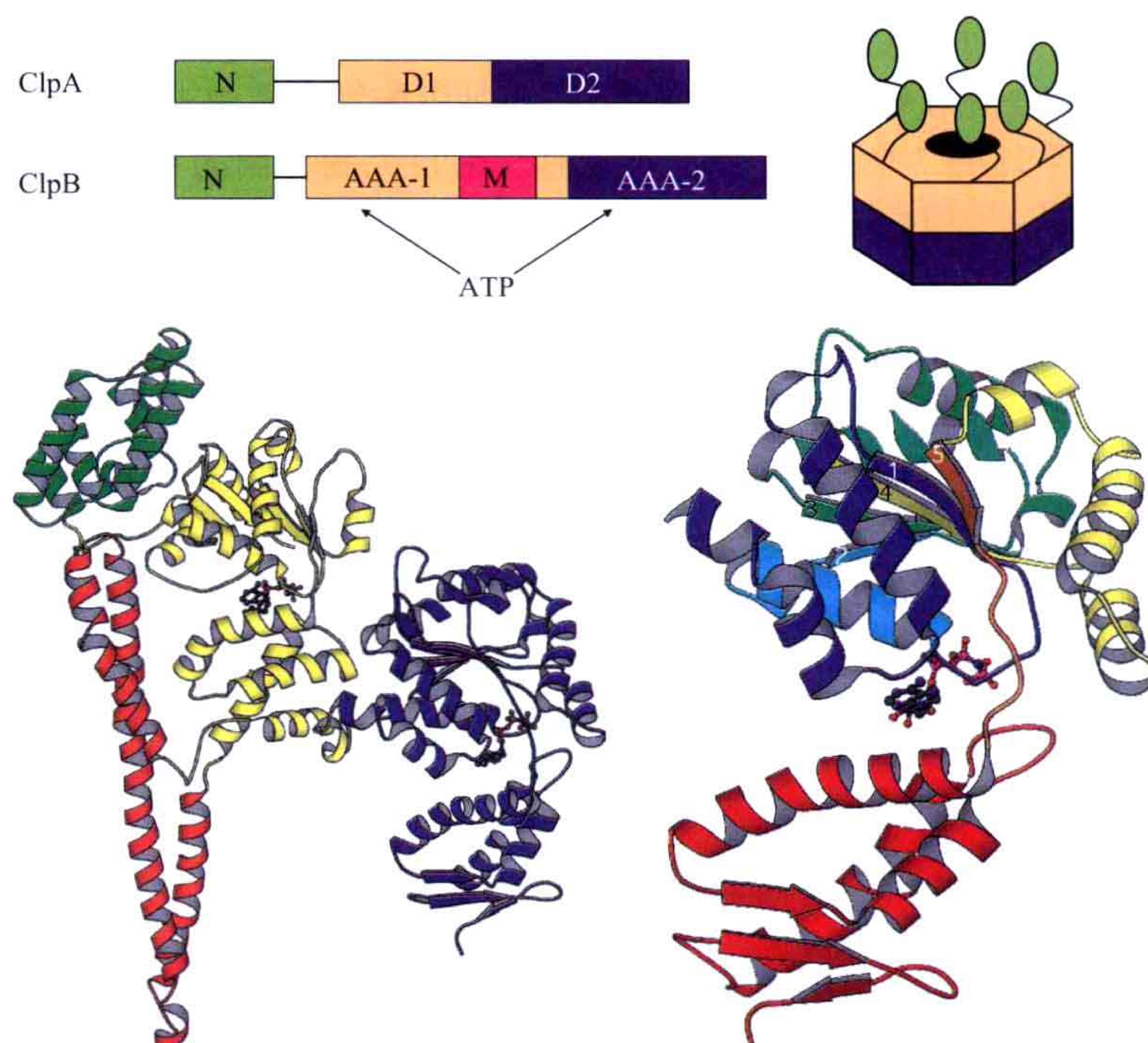


图 9.15　左上：Hsp100/ClpA 和 Hsp104/ClpB 的结构域图。右上：六聚体的 N 端结构域（绿色）通过一个柔性连接序列固定，AAA+ 结构域形成一个具有六重轴的双盘结构。左下：来自嗜热栖热菌（*T. thermophilus*）的 ClpB 单体结构。红色的中间（M）结构域在 Hsp100 中不存在，它在六聚体外表面形成类似于翅膀的结构。结构图中的颜色与左上结构域图的颜色相对应。ADPNP 分子以球棍模型表示（PDB: 1QVR）。右下：ClpB C 端的 AAA+ 结构域，其颜色自蓝（N 端）至红（C 端）显示。结合核苷酸的 P-loop 位于折叠片层中第一个折叠片之后。C 端螺旋束（红色）是所有 AAA+ ATP 酶结构域的一部分。

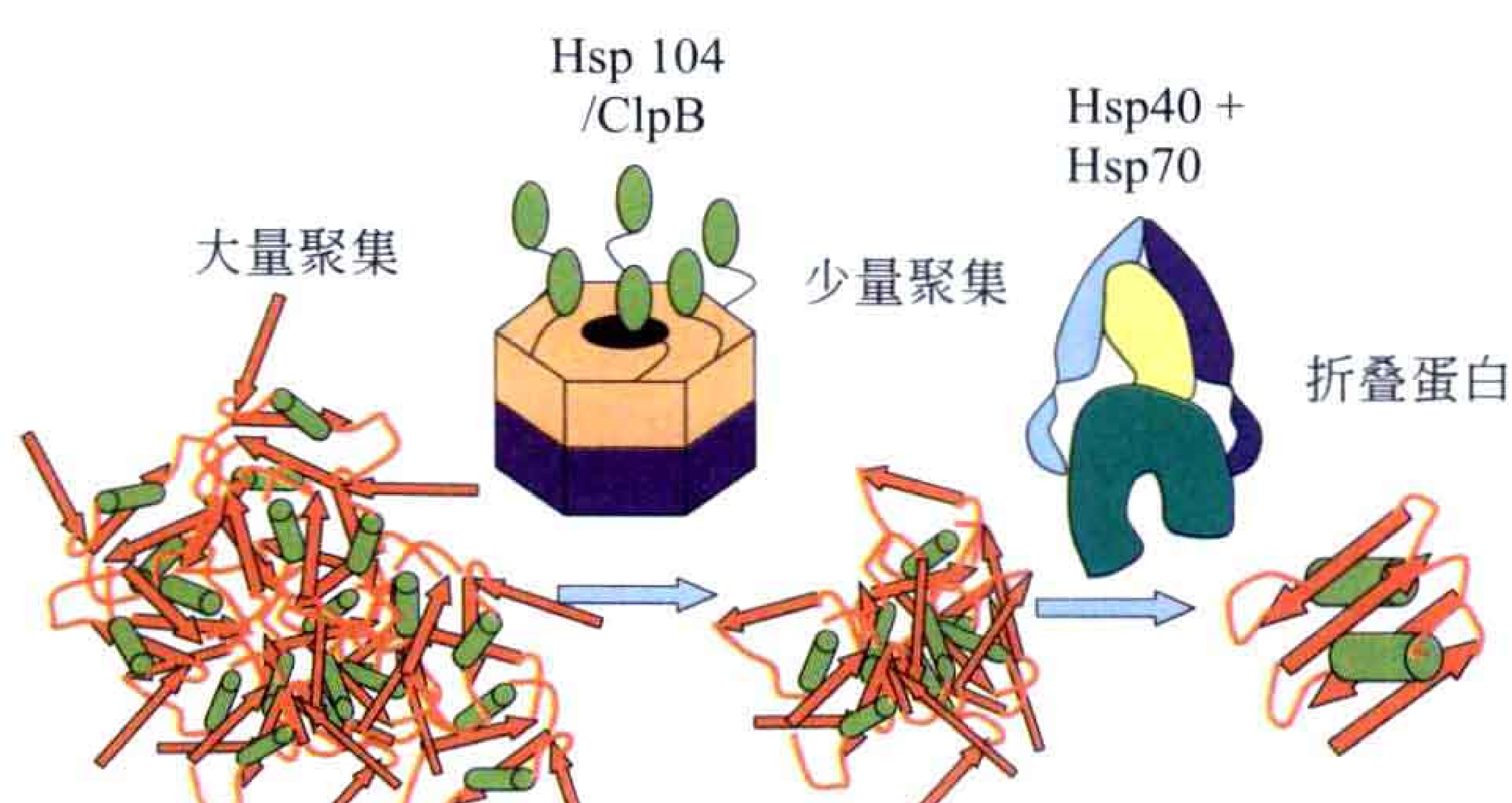


图 9.16　聚集体解聚的一条可能的途径。Hsp104/ClpB 把大的聚集体分解为小的聚集体，它们能作为 Hsp70/DnaK 家族分子伴侣的底物，形成天然折叠的蛋白质。

9.1.2.6　Chaperonin 类

被称为 chaperonin 的一类蛋白质以寡聚化的空腔形式存在，它们可捕获非天然折叠的蛋白质。其中最著名的是 GroEL，它属于 I 类 chaperonin。GroEL 是一个大家族蛋白中的一员。另一个 chaperonin 是热体（thermosome），它是一个古细菌蛋白，属于 II

类。I 类成员主要是同寡聚体，主要存在于细菌或内共生的细胞器（线粒体或叶绿体）中；II 类成员通常是异寡聚体，在古细菌或真核细胞的细胞质中被发现。chaperonin 由两个环组成，每个环有 7 个、8 个或 9 个亚基。来自大肠杆菌的 GroEL 每个环有 7 个亚基，而来自古细菌的热体和人的 TriC 每个环有 8 个亚基，来自芝田硫化叶菌（*Sulfolobus shibatae*）的 TF55 每个环有 9 个亚基。

GroEL chaperonin 也被称为 Hsp60，它是 GroE 蛋白复合体的一部分。这个复合体有两个部分，另外一部分是辅 chaperonin GroES 或 Hsp10。GroEL 由两个环组成，每个环有 7 个亚基。GroEL 环形成两个大空腔，其他蛋白质可结合于空腔内。GroES 是 7 亚基组成的盘状结构。GroEL 有三个结构域：赤道结构域（E）、中间结构域（I）和顶端结构域（A）。中间结构域位于其他两个结构域之间，发挥类似于铰链的功能。N 端和 C 端都位于 E 结构域中。能结合 ATP/ADP 的 E 结构域形成了两个环的作用界面，而 A 结构域位于 GroEL 圆柱体的两端（图 9.17）。

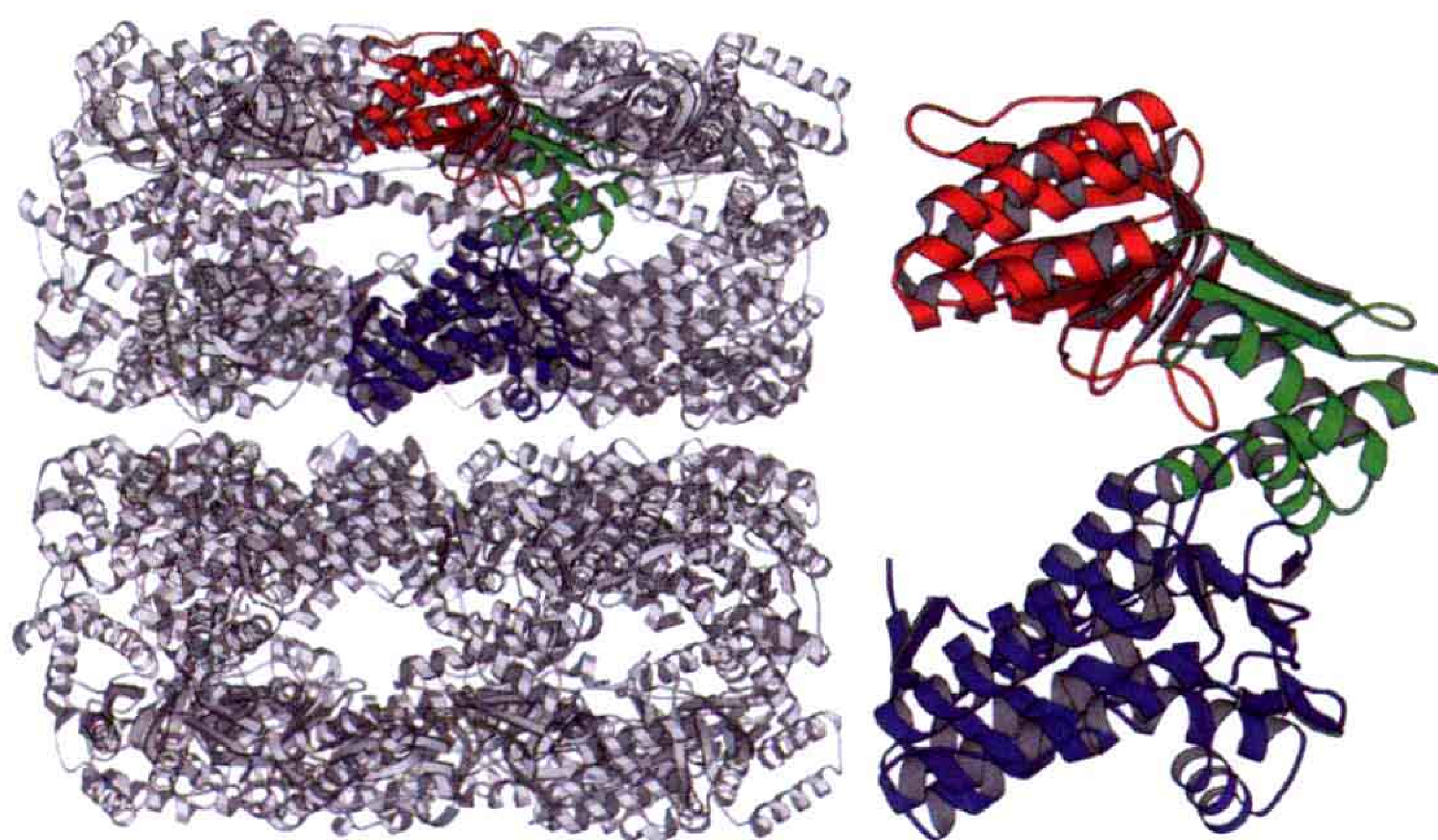


图 9.17　来自大肠杆菌的未结合核苷酸的 GroEL 结构。左侧显示的是两个七亚基圆环。14 个相同亚基中的一个亚基显示为彩色。右侧显示的是 GroEL 亚基的细致结构。赤道结构域（E，蓝色）、中间结构域（I，绿色）和顶端结构域（A，红色），如图所示（PDB: 1XCK）。

ATP 结合在一个亚基上可促进 ATP 与相同（顺式）环上的其他亚基结合（正协同），同时，阻止 ATP 与对面（反式）环上的亚基结合（负协同）。GroES 随后被激活并结合于顺式环上，但其与反式环的结合被阻遏（图 9.18）。

GroE 或其相关的 chaperonin 能帮助一系列广泛的底物进行折叠。顶端结构域可捕获未折叠或部分折叠的底物蛋白。与 GroES 结合的 GroEL 上的残基也会结合未正确折叠的蛋白质。以这种方式，底物蛋白被捕获于顺式环形成的大空腔中，并由 GroES 形成帽子。

结合有 ADP 或 ATP 核苷酸的 GroEL 可以结合 GroES。GroES 作为一个帽子结合于圆柱体的一端（图 9.18）。只有这个顺式环能结合核苷酸。与 GroES 的结合导致 GroEL 亚基发生巨大的构象变化。顶端结构域变得更加直立以便与 GroES 相互作用。结构域的构象重排如图 9.19 表示。在构象变化过程中，顶端结构域的七重对称形式因不同底物蛋白的结合可以产生扭曲，这在嗜热栖热菌的 GroEL-GroES 结构中被观察到（图 9.20）。

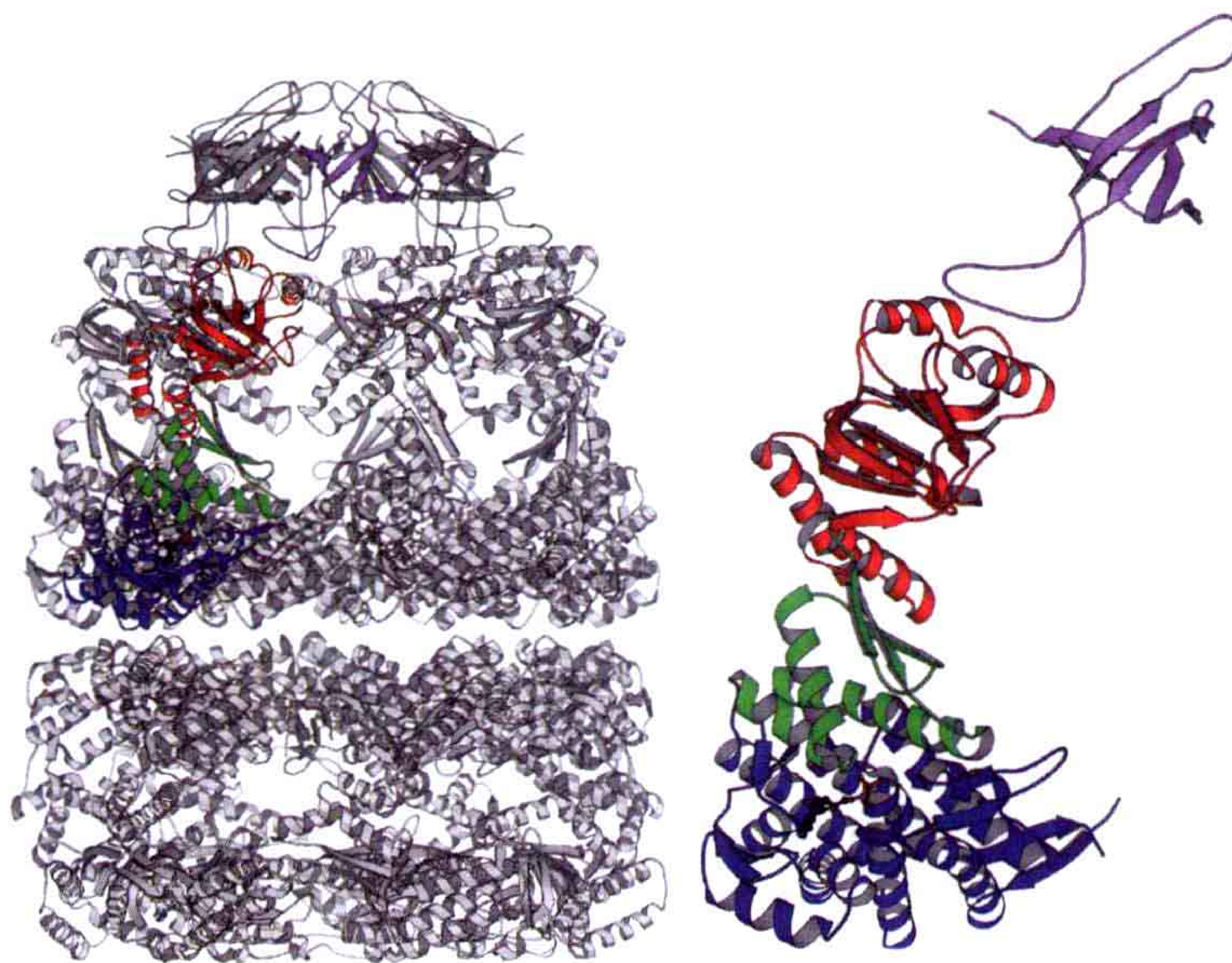


图 9.18　左：来自大肠杆菌的结合有 ADP 和 GroES 的 GroEL 结构。GroES（紫红色）以帽子的形式结合于圆柱体顺式环的一端，并改变了其结构。A 结构域（右）向上垂直移动以便和 GroES 相互作用。反式环几乎没有受影响。右：单一一个 GroES 和 GroEL 的结构（PDB: 1SX4）。

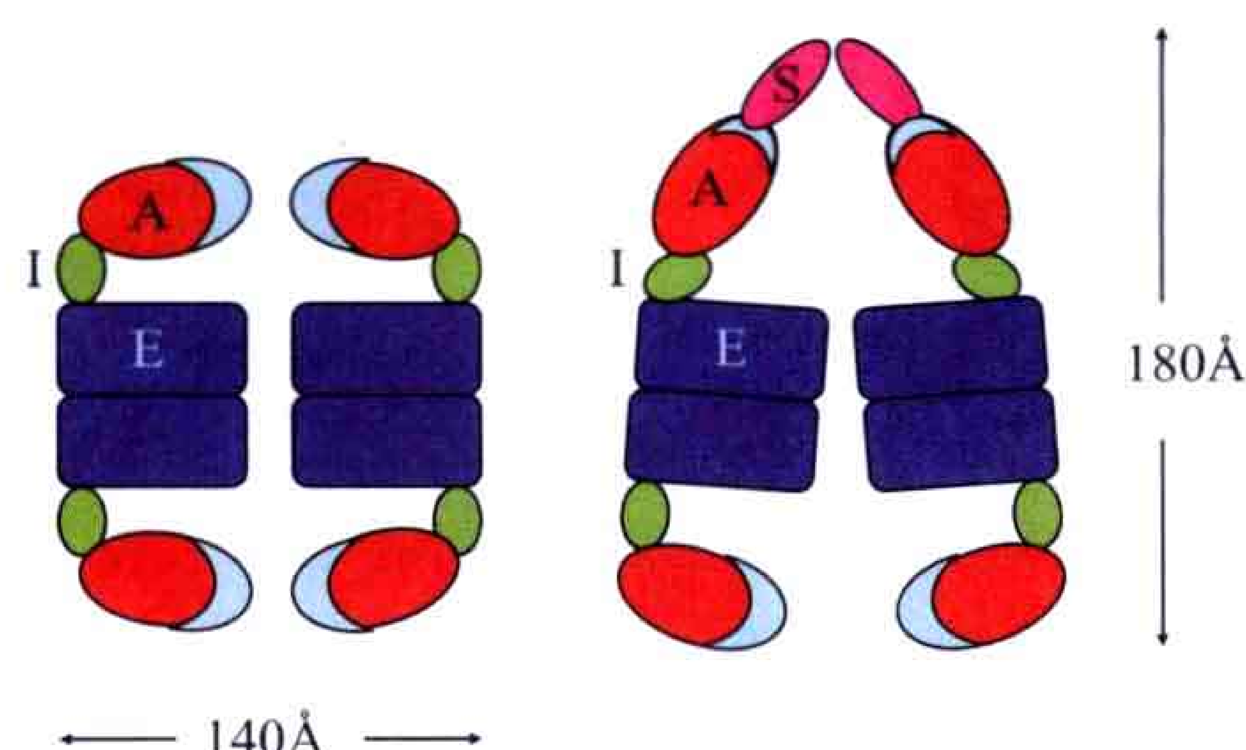


图 9.19　在未结合和结合 GroES 时 GroEL 结构域的重排。GroEL 的赤道结构域（E）、中间结构域（I）和顶端结构域的颜色及 GroES 的颜色与图 9.17 和图 9.18 相同。与 GroES 结合的 GroEL 七亚基圆环称为顺式环，相反的环称为反式环。在构象向左侧图示转变过程中，顶端结构域的疏水面（浅蓝色）暴露。这些疏水面在复合体中被 GroES 部分覆盖。

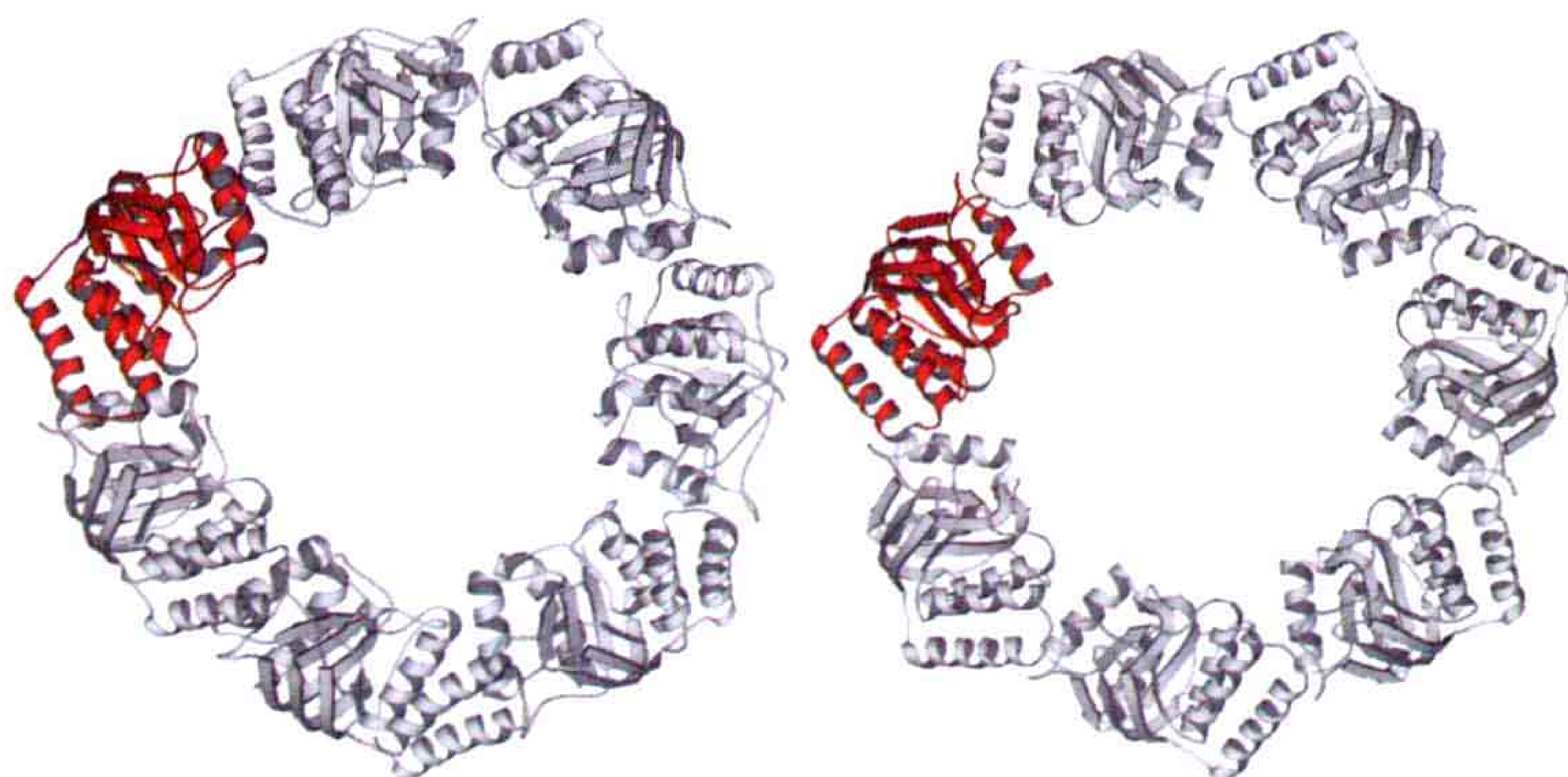


图 9.20　顺式环的顶端结构域不一定需要遵守七重对称。嗜热栖热菌的 GroEL-GroES 结构包含的 7 个亚基只是非常接近七重对称。各个亚基均可以存在两个稍微不同的朝向。这可能有赖于底物蛋白的结合（左，PDB:1WE3）。与在大肠杆菌中观察到的完美的七重对称（右）相比，此结构的偏差非常小。

GroEL 圆柱体的入口在特性上是疏水的，这可以使其与部分变性或非天然构象蛋白暴露出的疏水部分相互作用。结合 ATP 和 GroES 后产生的影响是顺式空腔变大，疏水残基隐藏于与 GroES 结构域接触或互作的部分。取而代之的是，主要带负电荷的亲水残基暴露于圆柱体的内部（图 9.21 和图 9.22）。亲水残基在物种之间非常保守，并可能对结合非天然构象的蛋白质非常重要。

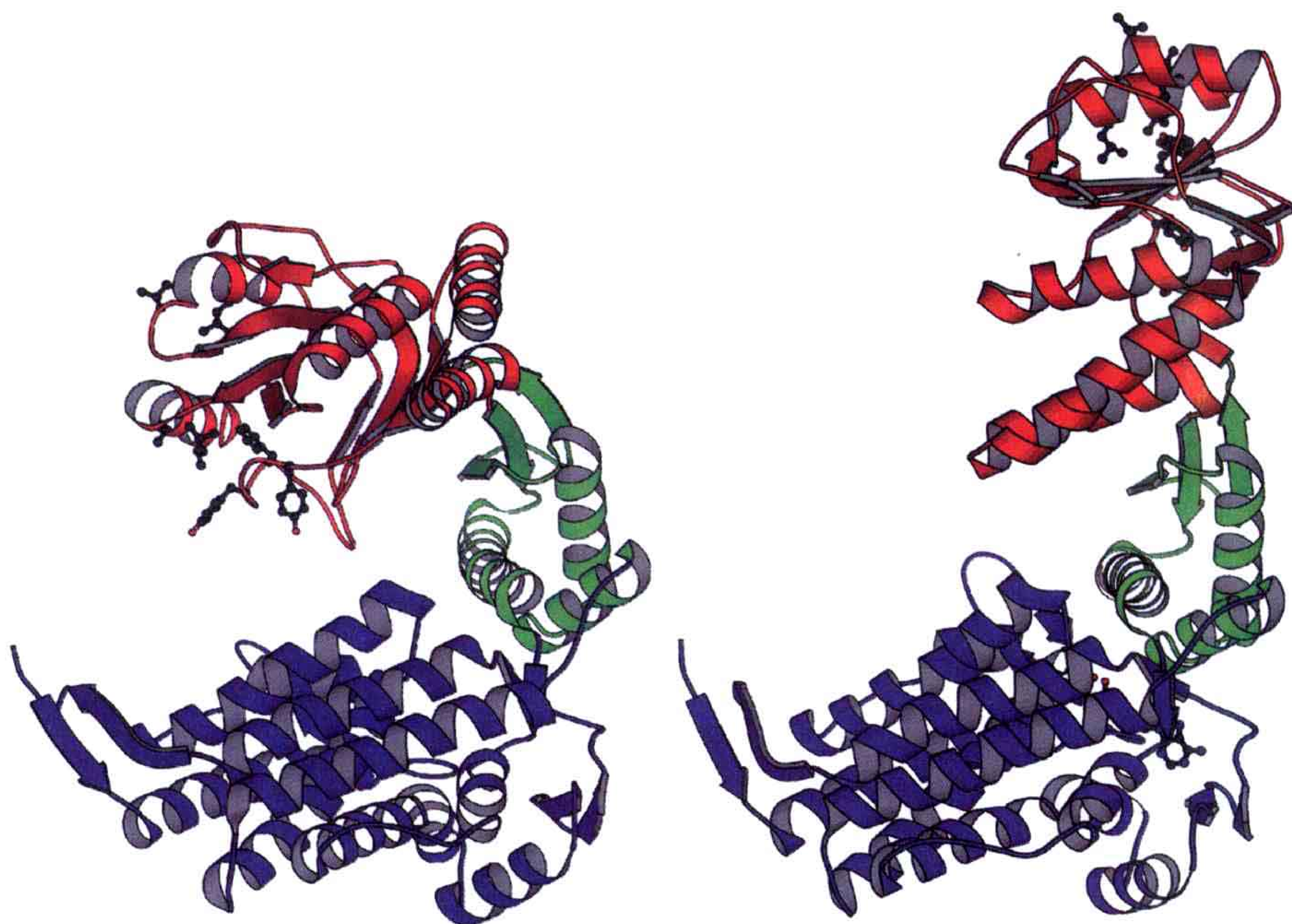


图 9.21 GroEL 亚基的两种构象。参与底物结合的疏水侧链以球棍模型表示。左：没有配体结合的 GroEL（PDB:1XCK）；右：与 ATP 和 GroES 结合的 GroEL（PDB:1SX4）。

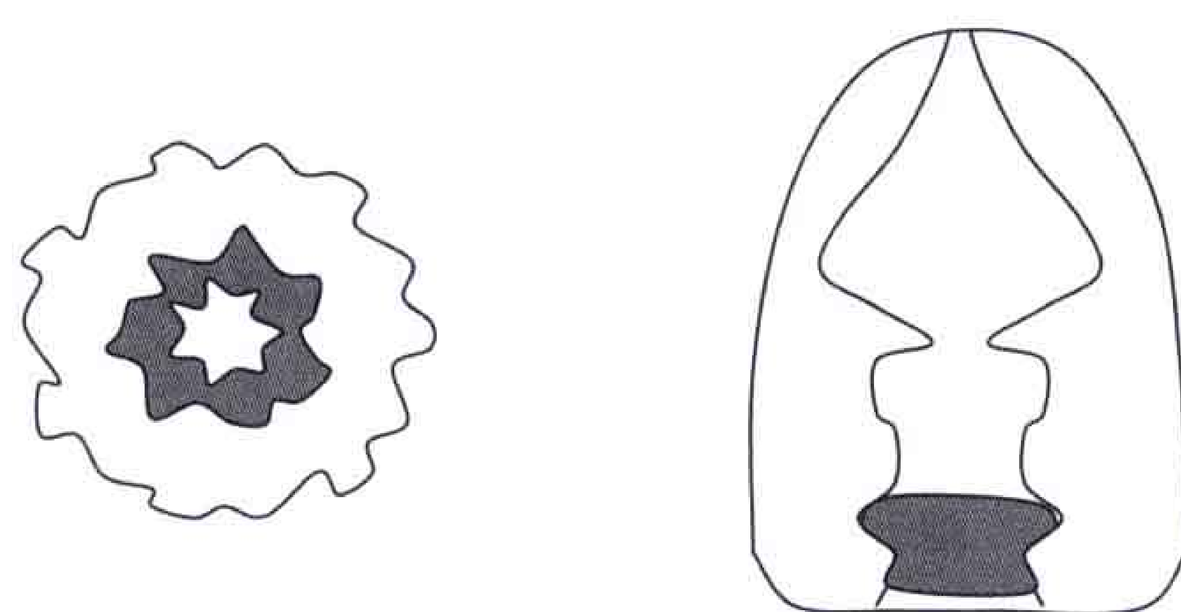


图 9.22 GroE 的疏水面（以灰色显示）。左：行使折叠作用的空腔开口处有一圈疏水氨基酸；右：GroEL-GroES 复合体。GroES 结合于顺式环的上方。下部的半个圆柱体，即反式环，有一个高度疏水的向内开口，准备结合非天然构象蛋白。上部的以 GroES 为帽子的半个圆柱已经变为扩大了的亲水空腔，未折叠的蛋白质被释放到这个亲水空腔内进行折叠。

因此，一个非天然折叠状态的蛋白质结合于顶端结构域的疏水氨基酸上，这些氨基酸位于 7 个 GroEL 亚基形成的圆柱体的入口处（图 9.23）。这个七亚基的环因此也就成为顺式环。A 结构域中疏水的部分（H 和 I 螺旋）是主要的底物接触位点。巨大的疏水面可能使引起结合的错误折叠的底物蛋白进一步去折叠。结合了底物的环将会结

合 ATP。这种结合是协同的：一旦一个亚基结合了 ATP，顺式环上的其他亚基也会被诱导结合 ATP。但是，GroEL 反式环上的亚基将产生负协同效应。随着核苷酸的结合，顶端结构域将产生有助于 GroES 帽子结合的构象变化。然后底物蛋白将经历一个环境的改变，从疏水相互作用到亲水相互作用并拥有一个更宽敞的空腔。这里没有可结合的疏水区，因此蛋白质有机会在空腔内进行自身的重折叠。现在，双圆柱机器中的反式圆柱已准备好了经历相同的步骤。当非天然构象的多肽和 ATP 分子结合于 GroES 的反式圆柱上时，已折叠好的多肽从前面的顺式环中释放，以前的反式环现在就变成了顺式环。

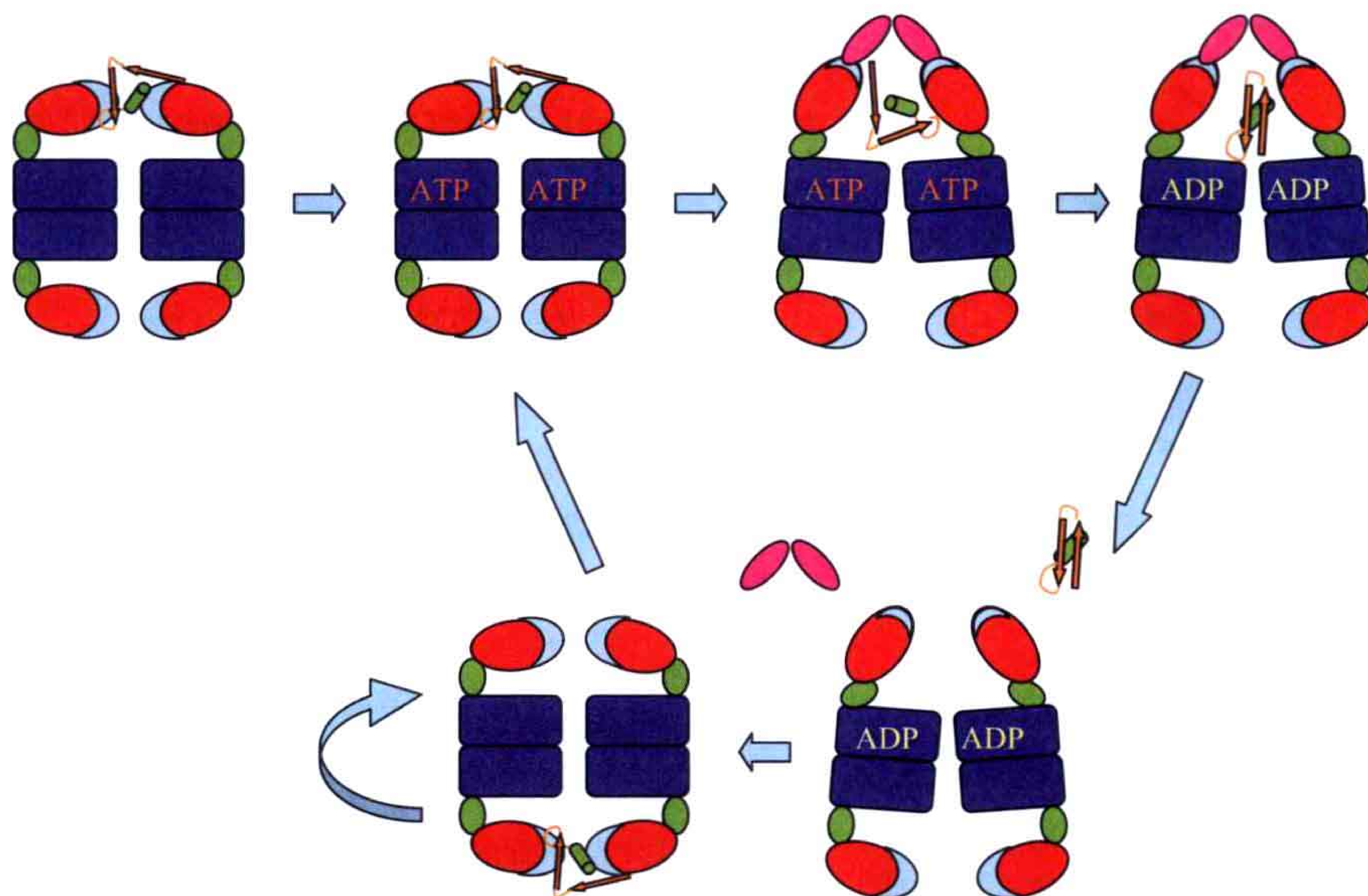


图 9.23　GroE 行使功能的循环过程。两个环之间在结合 ATP 上有负协同效应，但它们也顺序工作。当 GroES 结合于顺式环上时，反式环的开口扩大并能更容易地结合新的底物蛋白，以前的反式环就变成了新的顺式环。

在蛋白质折叠过程中，我们很难简单地给 GroE 确切的功能下一个定义。其功能可以是为蛋白质折叠创造一个合适的环境（一个极性的笼状结构），或者只是阻止那些还没有形成它们所具有的天然构象的蛋白质发生聚集（动力学分割）。通过与 GroE 结合的一次循环过程，一个还没有正确折叠的蛋白质可以重新进行折叠。GroE 也可以激活那些未折叠或错误折叠的蛋白质，让它们重折叠为天然构象。

9.1.3　蛋白质合成过程中新生肽链的折叠

细胞中新合成的蛋白质会暴露在拥挤的环境中。因此，新生肽链的正确折叠是翻译的一个重要方面。核糖体的出口孔道（见第八章）非常窄，不允许新生肽链进行有效地折叠。当一些蛋白质从出口孔道出来时，它们可能会同时折叠。但是很多情况下，

新生肽链的正确折叠需要分子伴侣。在原核和真核系统中，一些分子伴侣直接与核糖体和新生肽链相互作用。在细菌中，新生肽链主要与结合在核糖体上的小分子伴侣或“支持”分子伴侣相互作用。Trigger factor（TF）和 Hsp70（DnaK）属于这类分子伴侣。它们是单体蛋白，主要参与阻止正在合成的多肽发生聚集。

9.1.3.1 trigger factor

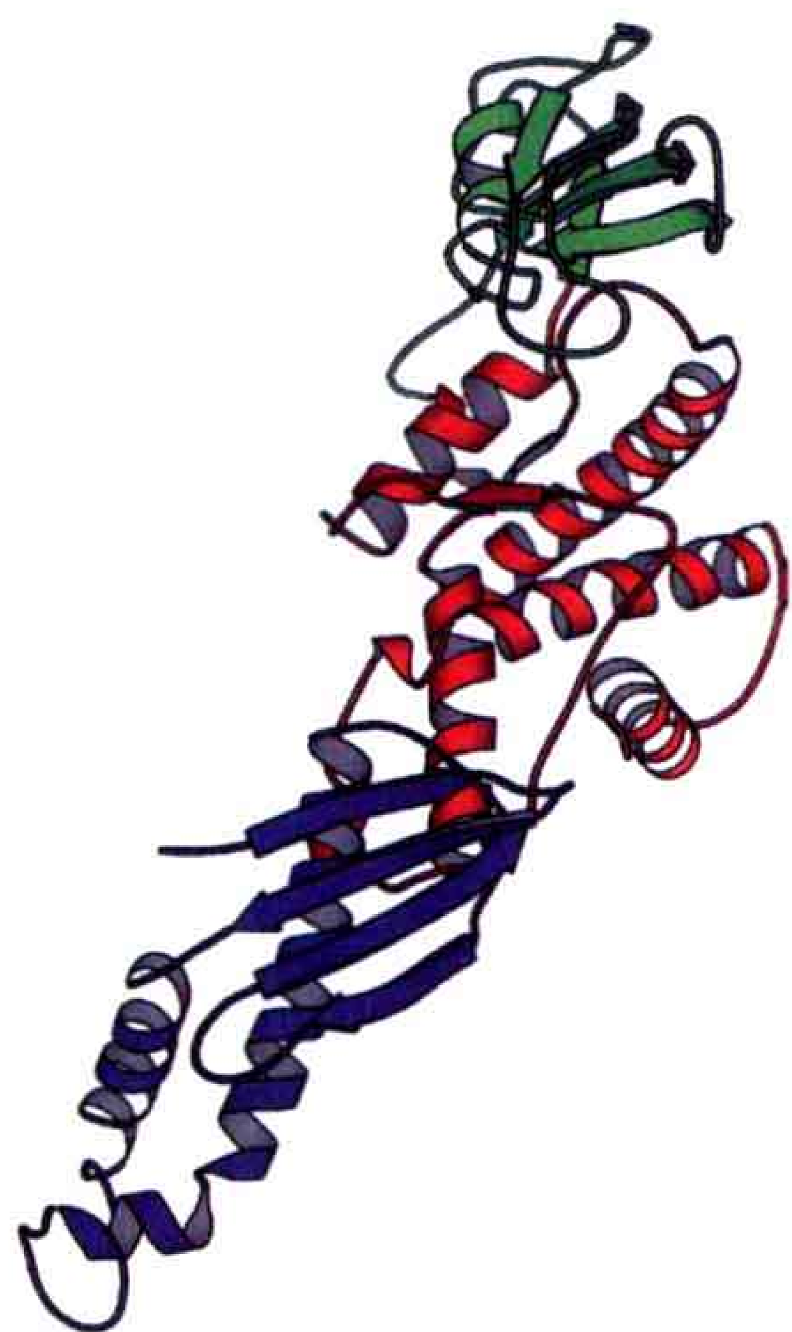


图 9.24 来自霍乱弧菌（*Vibrio cholerae*）的 TF 结构。N 端结构域，即核糖体结合结构域位于底部（蓝色），然后是 C 端结构域（红色），以及位于顶部的中间 PPlase 结构域（绿色）。底物结合位点在纵贯 TF 全长的一条沟中。

trigger factor（TF）是一个双功能蛋白，它通常情况下是第一个与新生肽链相互作用的非核糖体蛋白。它属于类 Hsp70 分子伴侣家族，但它同样也是一个肽基脯氨酰基顺反异构酶（PPIase）。TF 的一个核心功能是，它既能识别需要帮助折叠的新生肽，也能识别需要运出的肽。

TF 的结构已经被解析（图 9.24）。它由三个结构域组成：N 端结构域、中间结构域（PPIase）及 C 端结构域。核糖体结合位点和 PPIase 在分子上相反的两端。在不结合核糖体时，TF 沿蛋白质长轴形成二聚体。但是当结合核糖体时，TF 是一个单体。

TF 的 N 端结构域与核糖体大亚基结合的结构已经被解析。TF 的 N 端结构域结合于大亚基外表面出口通道的开口处，与核糖体蛋白 L23 和 L29 毗邻。

研究人员已经通过标记在 TF 上的发光探针分析了底物结合情况。不同长度的肽链已经经过分析。TF 在其纵贯全长的一条沟中结合底物。TF 的二聚面与肽链结合面位于同一位置。

在核糖体肽链出口通道中的及 TF 中的特定肽链的总长强有力地暗示肽链几乎完全伸展。新生多肽在其合成过程中，在 TF 内一步步前进，这一事实暗示了其与 TF 的结合力并不强。另外，新生多肽的结合区域位于一条开放的沟中。TF 可能提供了很好的保护，以免发生错误折叠和降解，但当足够可以形成蛋白质部分折叠的氨基酸结合于 TF 上时，它们将会从沟中脱离并完成折叠。

9.1.3.2 新生肽链的受助折叠

在细菌中，与 TF、Hsp40 和 Hsp70 的相互作用通常足以使蛋白质正确折叠（图 9.25）。在其他生物体中，蛋白质的正确折叠需要与更复杂的 chaperonin 相互作用，如 GroEL/GroES。据目前所了解的事实，chaperonin 不与核糖体相互作用。在真核生物中，完成折叠过程需要许多分子伴侣的参与。

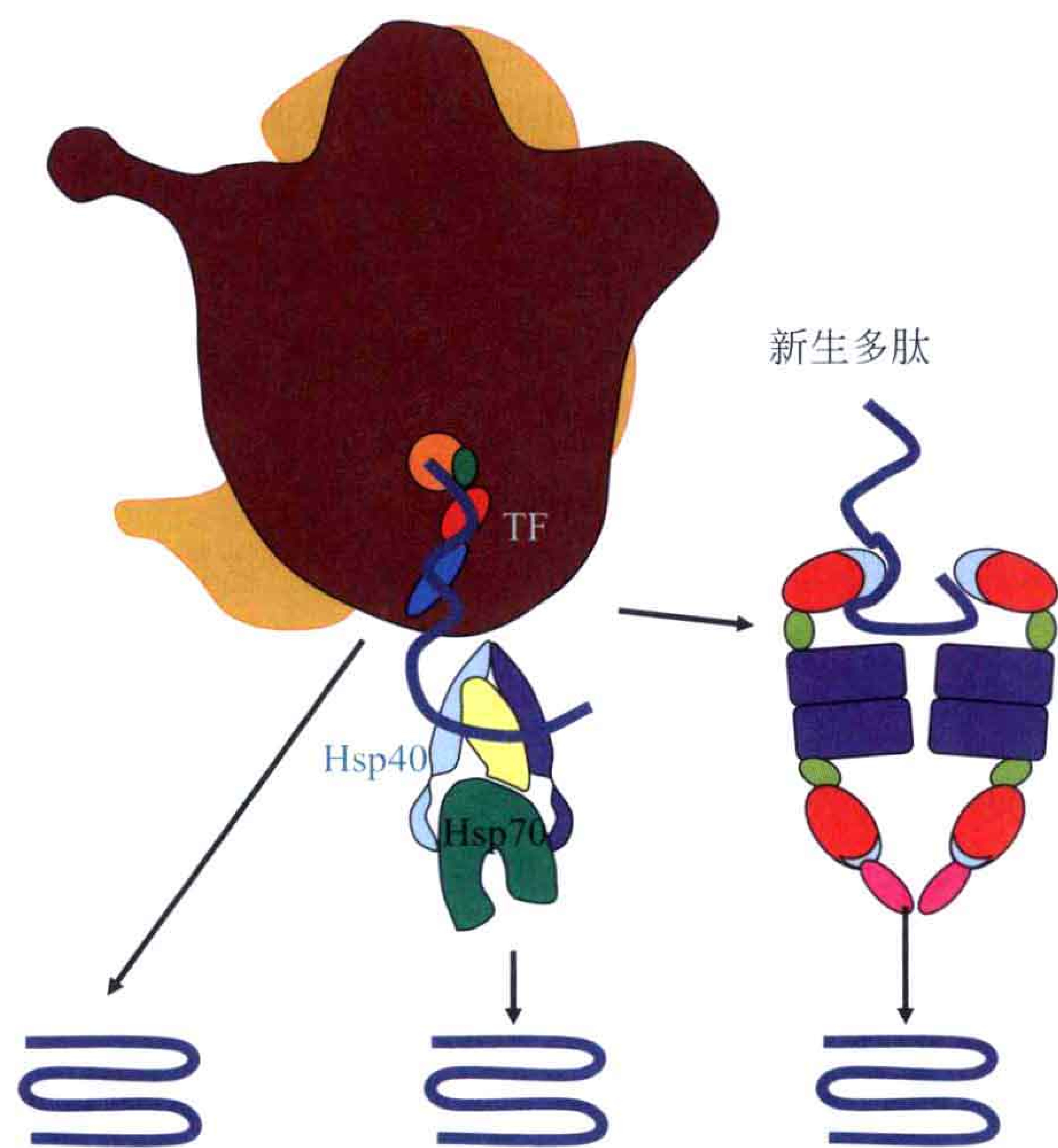


图 9.25　细菌中新生肽链的折叠。trigger factor（TF）是一个结合于核糖体上的分子伴侣。新生多肽的折叠需要 Hsp40 和 Hsp70，以及 GroEL/GroES 的帮助。

9.2　蛋白质降解

蛋白质的周转是生命体的核心特征。蛋白质的降解可能和它们的合成同样重要。一个活细胞需要持续提供氨基酸以合成新的蛋白质。这些氨基酸可以来自饮食中蛋白质的降解，也可以来自细胞蛋白的降解。因此不论胞内还是胞外，蛋白质均能被水解。真核细胞中存在一种称为溶酶体的细胞器，可降解包括蛋白质在内的大分子。饮食中蛋白质的降解大部分在胃肠道内通过消化系统中的胞外蛋白水解酶完成，如胃蛋白酶、胰蛋白酶、糜蛋白酶（图 1.2）。在低等生物体中，类似的需求通过胞外蛋白酶提供，通过降解外部的蛋白质来供给氨基酸。因为这些蛋白酶在各类生化教材中均有介绍，在这里只简要列出（表 9.2）。

表 9.2　一些单体蛋白酶家族

蛋白酶类型	活性位点残基	序列模体	例子
Ser（胰蛋白酶类）	Ser, His, Asp	—	胰蛋白酶
Ser（枯草芽孢杆菌蛋白酶类）	Asp, His, Ser	—	枯草芽孢杆菌蛋白酶
Cys	Cys, His	—	木瓜蛋白酶
Acid	Asp, Asp	—	胃蛋白酶
Zn	Zn 配体：His, His, Glu	HEXXH	羧肽酶 A

注：进一步的信息参见 MEROPS 数据库。

细胞中蛋白质的寿命各不相同，这与它们的功能有关。一些蛋白质寿命很长，而一些蛋白质可能只有几分钟的寿命。不同的蛋白质水解活性控制一个蛋白质的寿命。

胞内蛋白酶有两个主要的控制功能：调控和质量控制。蛋白质水解必须特异且必须完全被控制，以免随机破坏细胞中的蛋白质。蛋白质水解功能有如下特点：

（i）受那些需要特异性激活的酶类调控，如在血凝系统中。

（ii）被各种细胞中的蛋白抑制剂抑制。

（iii）不能接近那些正确折叠的蛋白质。

细胞中蛋白质的降解由非特异性的蛋白酶完成，它们中的绝大多数是寡聚体。这些蛋白酶能控制其活性，只使某一类底物被降解。通过结构的寡聚化，其活性位点不能接近折叠的蛋白质。

一些需要降解的蛋白质会被标记上泛素。这些标记的蛋白质随后被带入蛋白酶体，在这里它们被降解为片段，这些片段随后被多种酶降解为氨基酸。

在受到感染时，被 MHC 分子（见第十四章）呈递的抗原肽由蛋白酶体产生。这些外源肽段发出了细胞被感染和应该被销毁的信号。

非天然折叠的蛋白质可带来严重的健康问题，因为它们可能聚集（2.2.3.4 节和 9.1.2 节）。依靠分子伴侣的重折叠和被水解代表了两个处理错误折叠的蛋白质的主要路线。

许多寡聚的胞内蛋白酶是 ATP 酶，但也有一大类是不需要 ATP 的。ATP 常用于蛋白质降解前的去折叠过程。这一功能由一类常见的结构域或亚结构单位完成，即 AAA+ 结构域，它也在一些分子伴侣内被发现，这些分子伴侣可以使错误折叠的蛋白质重折叠或去聚集。

9.2.1 Serpin：蛋白酶抑制剂

有一大类蛋白质作为蛋白酶抑制剂发挥作用。其中丝氨酸蛋白酶抑制剂有 20 多个家族，在高等生物中最丰富的丝氨酸蛋白酶抑制剂是 serpin。

serpin 是一大类丝氨酸蛋白酶抑制剂家族，发现于所有的生命分枝中，包括细菌、古细菌、真核生物及病毒。它们是自杀性抑制剂，与其特异的抑制靶标以定比的方式被消耗。这个家族的一些成员包括：抗胰蛋白酶因子（antitrypsin）、抗糜蛋白酶因子（antichymotrypsin, ACT）、抗凝血酶因子（antithrombin）、纤溶酶原激活抑制因子 -1（plasminogen activator inhibitor-1, PAI-1）、α1- 蛋白酶抑制因子（α1-proteinase inhibitor, α1-PI）及卵清蛋白。卵清蛋白在蛋清中占蛋白质的 2/3，它有与 serpin 大体相同的结构，但它不是作为一种抑制剂发挥作用。serpin 参与在所有人体内受到严谨调控的蛋白质水解途径中：血凝反应、纤溶反应、补体级联反应、组织重塑、肿瘤转移、炎症及凋亡途径。尽管 serpin 之间序列同源性很低，但它们的结构非常相似，有 9 个 α 螺旋（A ~ I）和 3 个 β 折叠（A ~ C）。

这些 serpin 家族蛋白起初折叠为一个亚稳态结构，通过蛋白质水解可转化为一个稳定结构。它们的这个特征起初被认为是特有的，不过现在又知道了很多其他亚稳态的

蛋白质。起初它们折叠为一个亚稳态结构，这个亚稳态结构可通过共价的或非共价的互作转换为一个稳态的形式。另一个亚稳态蛋白的例子是朊蛋白，它们有一个可以转变为致病性稳态形式的正常亚稳态结构。

天然的 serpin 在动力学上滞留在一个高能状态，这个高能状态可以通过下面所示的两种方式被释放。典型的 serpin 亚稳态结构（紧张态）的熔解温度是 60℃，而稳态形式（松弛态）的熔解温度是 120℃。在亚稳态形式中，20~25 个氨基酸的反应中心卷曲（RCL）处于暴露状态，且具有非常大的柔性。这个卷曲可以以一种不同寻常的方式并入结构中：它以一条新的 β 折叠片形式插入 β 片层 A 中间（图 9.26）。这种类型的结构重排是不曾预料到的，但现在其性质已经被详细描述。

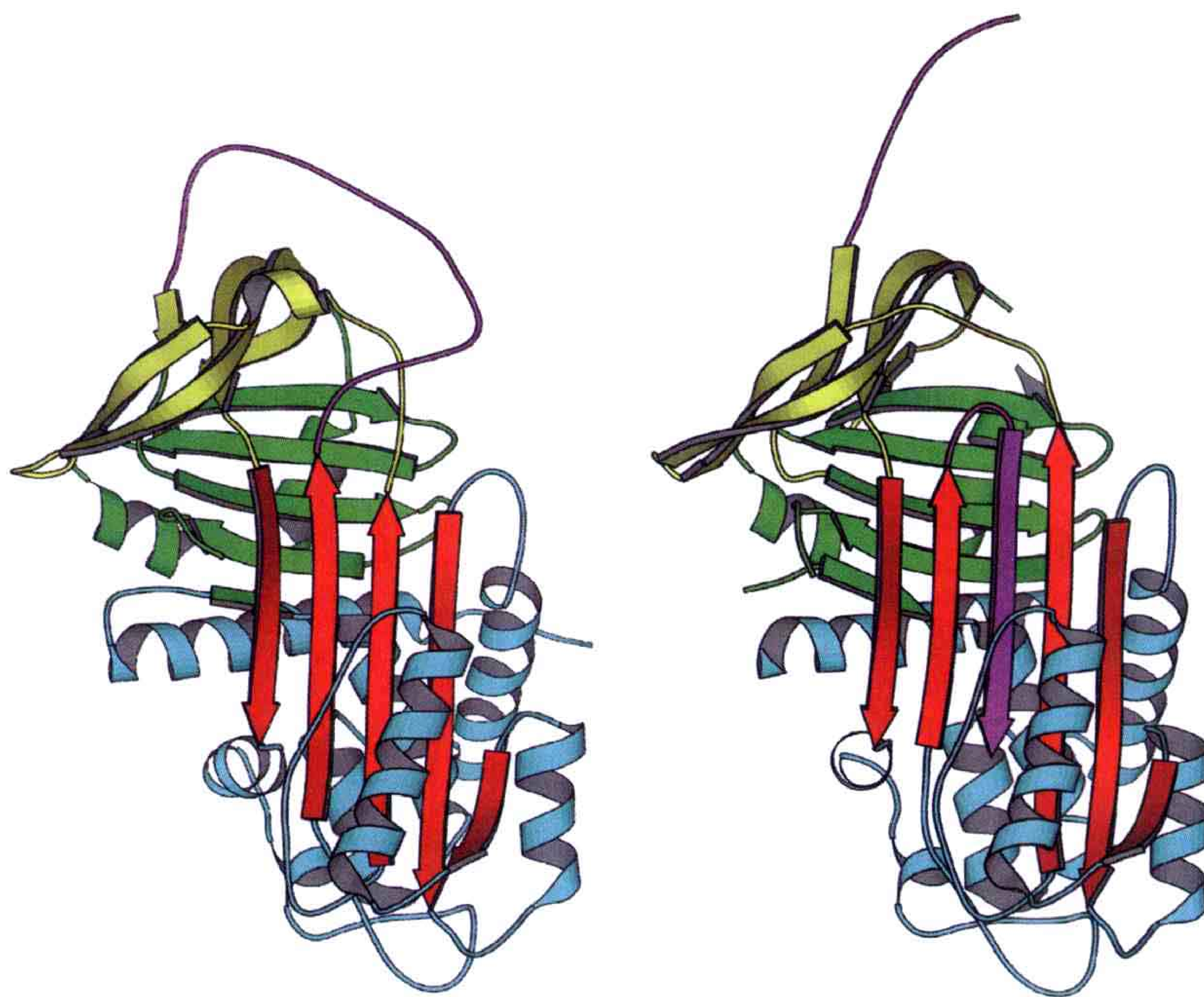


图 9.26　抗胰蛋白酶因子的亚稳态或紧张态结构（左，PDB: 1HP7）及稳定的松弛态结构（右，PDB: 1QMB）。处于亚稳态的反应中心卷曲（RCL，紫色）被剪切后并入片层 A 中间，并转变为稳定的松弛态。

RCL 卷曲的主要功能是其丝氨酸蛋白酶抑制剂活性。不同的蛋白酶识别卷曲上特定的氨基酸序列并将其作为底物（图 9.27）。它们结合于卷曲上并催化卷曲的断裂。然而，酶活性位点的丝氨酸仍共价连接于 serpin 切割位点的 N 端。这导致了 serpin 产生了一个巨大的构象变化，即卷曲插入了 β 片层 A 的中间。这意味着蛋白酶从 serpin 的近远端发生了大约 75Å 的移位。在这个状态下的复合体晶体结构显示：蛋白酶的一些部分柔性大大增加，并由晶体学中的 B 因子证实（图 9.27）。通过 NMR 方法分析胰蛋白酶和抗胰蛋白酶因子复合体表明：蛋白酶在这个状态下呈现为一个熔融的小球。蛋白酶中的柔性区域现在将易受酶解，并将导致其降解。如果 serpin 中的 RCL 已经被切割，那么将不能再被利用，因此它将被降解。

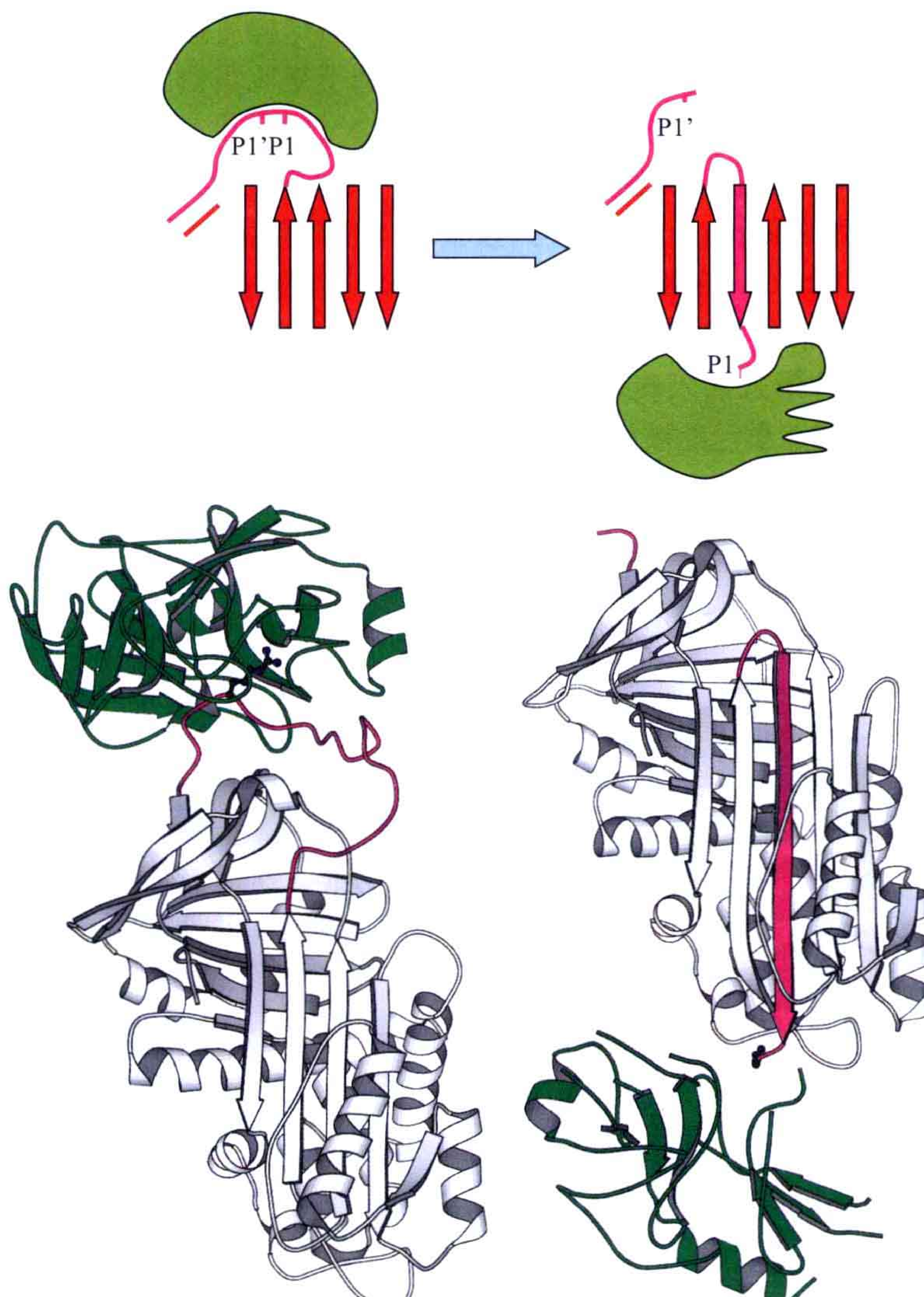


图 9.27 serpin 行为的简化图（上）和原理图（下）。左：蛋白酶（绿色）结合于 serpin 的 RCL 卷曲（紫色）上，卷曲连在片层中心的一个 β 折叠上（PDB: 1OPH）。右：这导致了 RCL 卷曲的水解，并以一个 β 折叠的形式并入 serpin 的片层 A 中（PDB: 1EZX）。作为这个过程的一部分，蛋白酶摆动到 serpin 的另一端。相互作用导致蛋白酶具有柔性，这使其容易受到酶解进而被除去。

serpin 在 RCL 不被切割时也可经历 β 折叠插入过程。这是抑制因子转化为潜伏的、非活性状态的一个调节机制。图 9.28 显示了在肽段不被切割时，RCL 卷曲插入 β 片层 A 中心时的一些中间步骤结构。另外，serpin 也能寡聚化。RCL 插入其他 serpin 分子的 A 片层中仅仅是多种可能的寡聚方式中的一种。

显而易见，亚稳定状态有一个与高度稳定状态的蛋白质不相容的结构元件。这些元件包括暴露的疏水基团及被包埋的没有形成合适氢键的亲水基团。同时，RCL 包含能与 β 片层 A 中间的折叠极好匹配的结构元件。甚至 RCL 部分插入片层中也是能量上有利的一步。

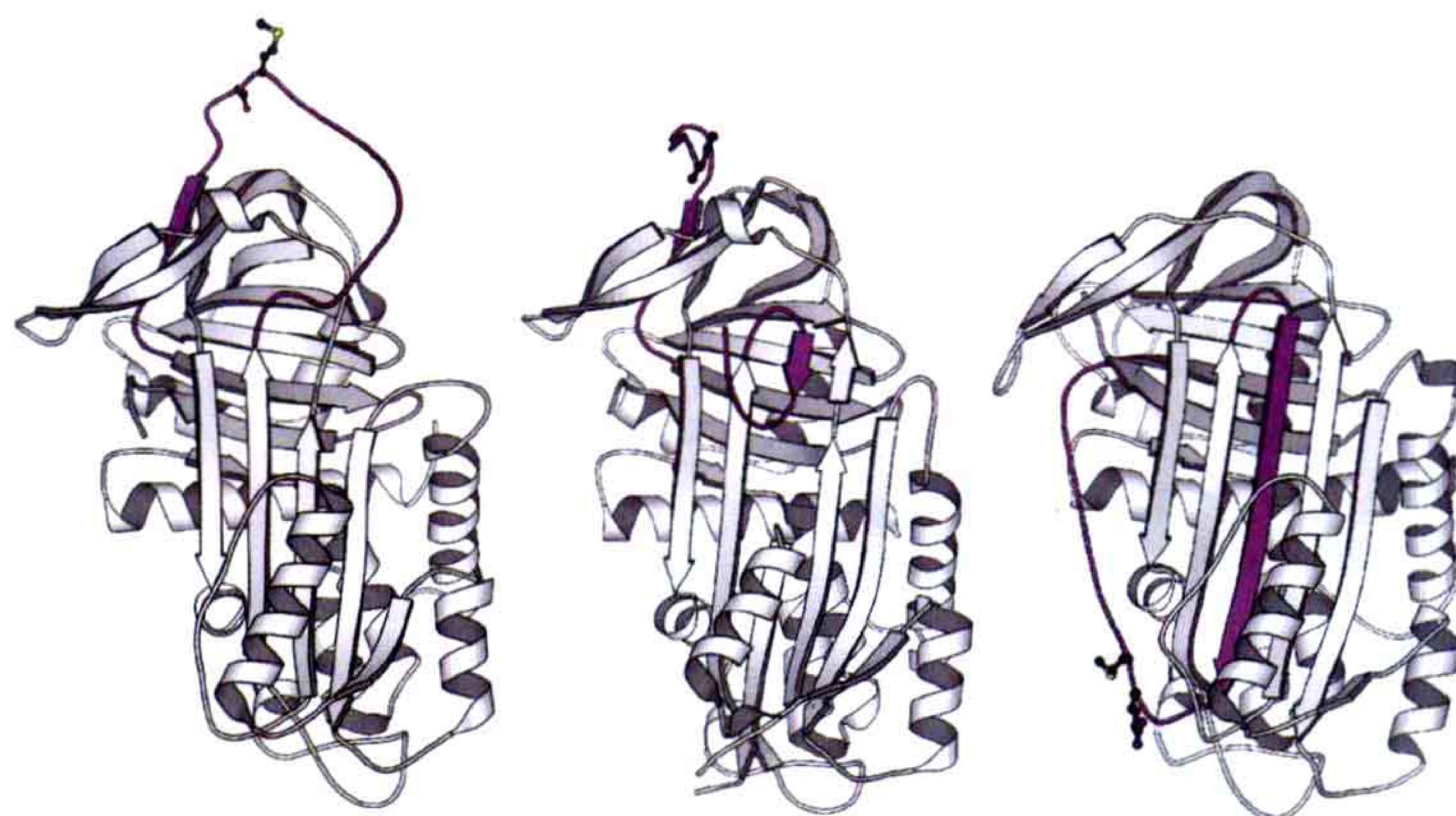


图 9.28　RCL 卷曲（紫色）逐渐的、自发的并入 β 片层 A 中心的结构图。卷曲中的肽键并未被切断，但抑制因子转化为一种潜伏的、非活性状态的形式。在活性形式的抑制因子中，将被切断的肽键两侧的氨基酸侧链以球棍模型显示。从左开始：α- 抗胰蛋白酶因子（PDB: 1QLP）、δ 形式的 α- 抗糜蛋白酶因子（PDB: 1QMN），以及纤溶酶原激活抑制因子 PAI-1（PDB: 1LJ5）。

9.2.2　泛素途径和蛋白酶体

真核细胞中的两个主要蛋白质水解系统是溶酶体和蛋白酶体。溶酶体是一种包含许多降解酶的细胞器。蛋白酶体在细胞内负责把蛋白质降解为小片段的寡聚体，这些小片段被更小的蛋白酶降解为氨基酸。蛋白酶体降解那些被共价连接的泛素标记了的蛋白质。这个系统被称为“垃圾处理”系统。

9.2.2.1　泛素标记系统

泛素是一个小蛋白，有 76 个氨基酸。它能共价连接到待降解的蛋白质上。它在 C 端含有一个 Gly-Gly 序列，这是与目的蛋白连接的位点。有三种蛋白质（E1、 E2、E3）参与标记过程（图 9.29）。泛素激活蛋白（E1）先将泛素 C 端腺苷酸化，随后将泛素转移到 E1 上并与泛素之间形成硫酯键。然后，泛素被转移到泛素接合酶（E2）的半胱氨酸上。最后，泛素 - 蛋白连接酶（E3）催化底物蛋白中的一个赖氨酸残基和泛素 C 端的甘氨酸之间形成异肽键，将泛素共价连接到底物蛋白上。虽然哺乳动物细胞中只有一种类型的 E1 蛋白，但是存在大约 50 种 E2 和大约 1000 种 E3 酶。E3 分子对不同类型的靶蛋白具有特异性，并可以分为三种不同类型。E2 接合分子对不同类型的 E3 具有特异性。E1 最终可以将泛素转移至所有 E2 上。

在细胞中，泛素标记具有多种不同的功能。蛋白酶体参与的降解信号需要至少 4 个泛素分子提供（图 9.29）。底物蛋白上也可以连接很多类似于泛素的蛋白质，但它们有其他的细胞功能，而不是作为降解的标记。SUMO（small ubiquitin-related modifier）可能是最受关注的类似于泛素的蛋白质，因为它可以标记不同的转录因子。

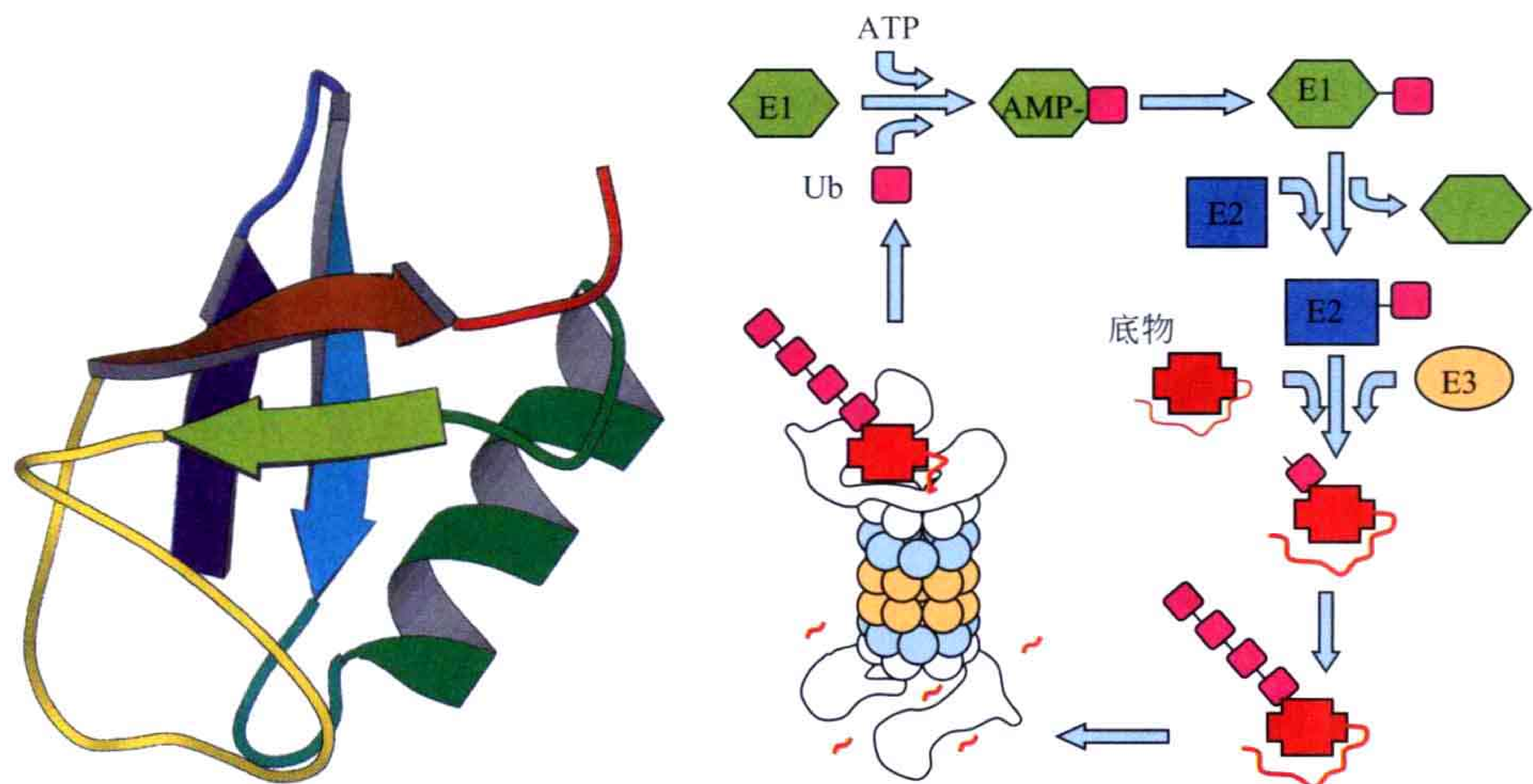


图 9.29　左：泛素分子（PDB: 1UBQ）；右：泛素标记底物蛋白（红色）的步骤。E1（绿色）是泛素激活蛋白。它首先将 ATP 分子中的 AMP 部分转移至泛素分子上来激活泛素，再将泛素转移至 E1 的一个半胱氨酸残基上。泛素接合酶（E2，蓝色）随后与泛素连接。最后，泛素连接酶（E3，浅棕色）将泛素转移至底物蛋白的一个赖氨酸残基的氨基上。待降解的底物变为多泛素化。多泛素化的蛋白质与蛋白酶体结合并去折叠，这导致底物蛋白转移至具有水解作用的空腔中并随之降解。泛素分子被释放并重新利用。

泛素有 7 个赖氨酸残基，其中 4 个赖氨酸（K11、K29、K48、K63）参与到和多聚泛素中下一个泛素分子的 C 端甘氨酸形成异肽键的过程中。通过 K48 和 K29 连接的多聚泛素是蛋白酶体识别的降解信号，而靶蛋白的单泛素化具有不同的细胞功能。

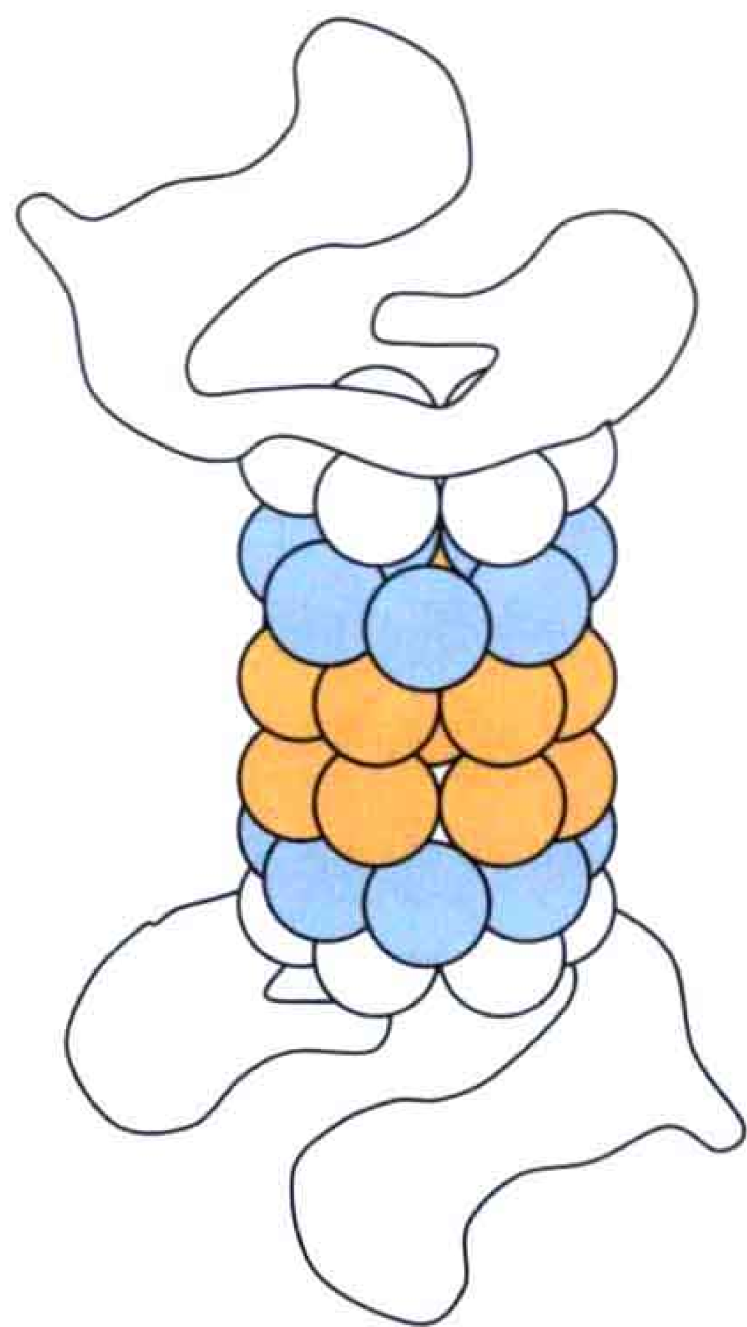


图 9.30　26S 蛋白酶体的示意图。中间的 20S 核心（有色的部分）由 4 个环组成，每个环有 7 个亚基。这个核心结构的性质已经被详尽地了解。19S 部分（白色）从两端延伸出来，其性质并未被完全了解。19S 的基座是由 6 个相同亚基组成的环状结构。

9.2.3　蛋白酶体

蛋白酶体在所有生命分枝中均被发现。它们虽然对于真核细胞来说是必需的，但对于古细菌的存活并不必要，并且只在部分真细菌中被发现。蛋白酶体是细胞内多泛素化蛋白受调控地降解为寡肽的地点。底物可以是短寿命的调节蛋白或非天然折叠状态的蛋白质。蛋白酶体降解后产生的肽段长度一般是 7~9 个残基。这些肽段可以被其他的蛋白酶进一步处理。

完整的蛋白酶体是一个 26S 的颗粒。总分子质量是 2.5 MDa。真核生物的 26S 蛋白酶体由大约 30 个不同的亚基组成。图 9.30 显示的是通过冷冻电镜断层照相获得的低分辨率 26S 蛋白酶体的图像轮

廓。两个 19S 部分从中间结构中延伸出来，这个中间结构即核心。该核心即 20S 蛋白酶体（700 kDa），是一个由 4 个七元环组成的圆柱体，高 150Å，直径为 110Å。这 4 个环每个环有 7 个亚基。外侧的两个环由 α 亚基组成，内侧的两个环以头对头的方向排列，由 14 个 β 亚基组成。

两种类型的亚基有一个共同的折叠形式，即两个 5 股的 β 片层及片层两侧每侧各两个 α 螺旋组成的三明治结构（图 9.31）。H1 和 H2 螺旋负责 α 和 β 亚基之间的接触，而 H3 和 H4 螺旋构成了 β 环之间的接触。所有的 20S 蛋白酶体具有相同的基本结构排布。β 亚基组成的环形成了一个空腔，而 α 亚基在空腔的入口处形成了一个窄的环状结构。催化酶解的活性位点位于 β 亚基的内侧（图 9.32）。因此，一个蛋白酶体内可以有 14 个催化酶解的活性位点。

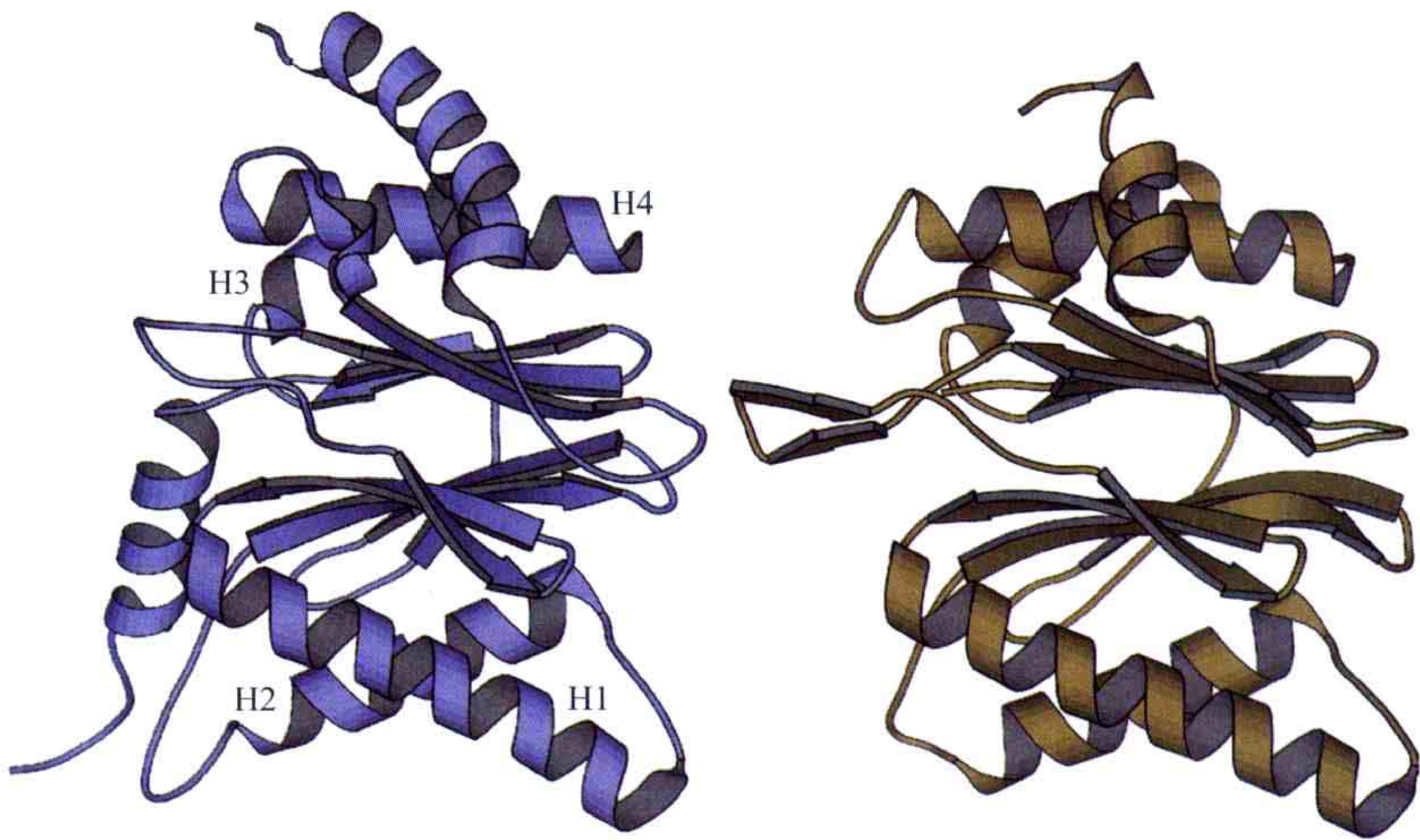


图 9.31　来自古细菌闪烁古生球菌（*Archaeoglobus fulgidus*）的蛋白酶体 α 亚基（左）和 β 亚基（右）具有非常相似的结构（PDB: 1J2Q）。

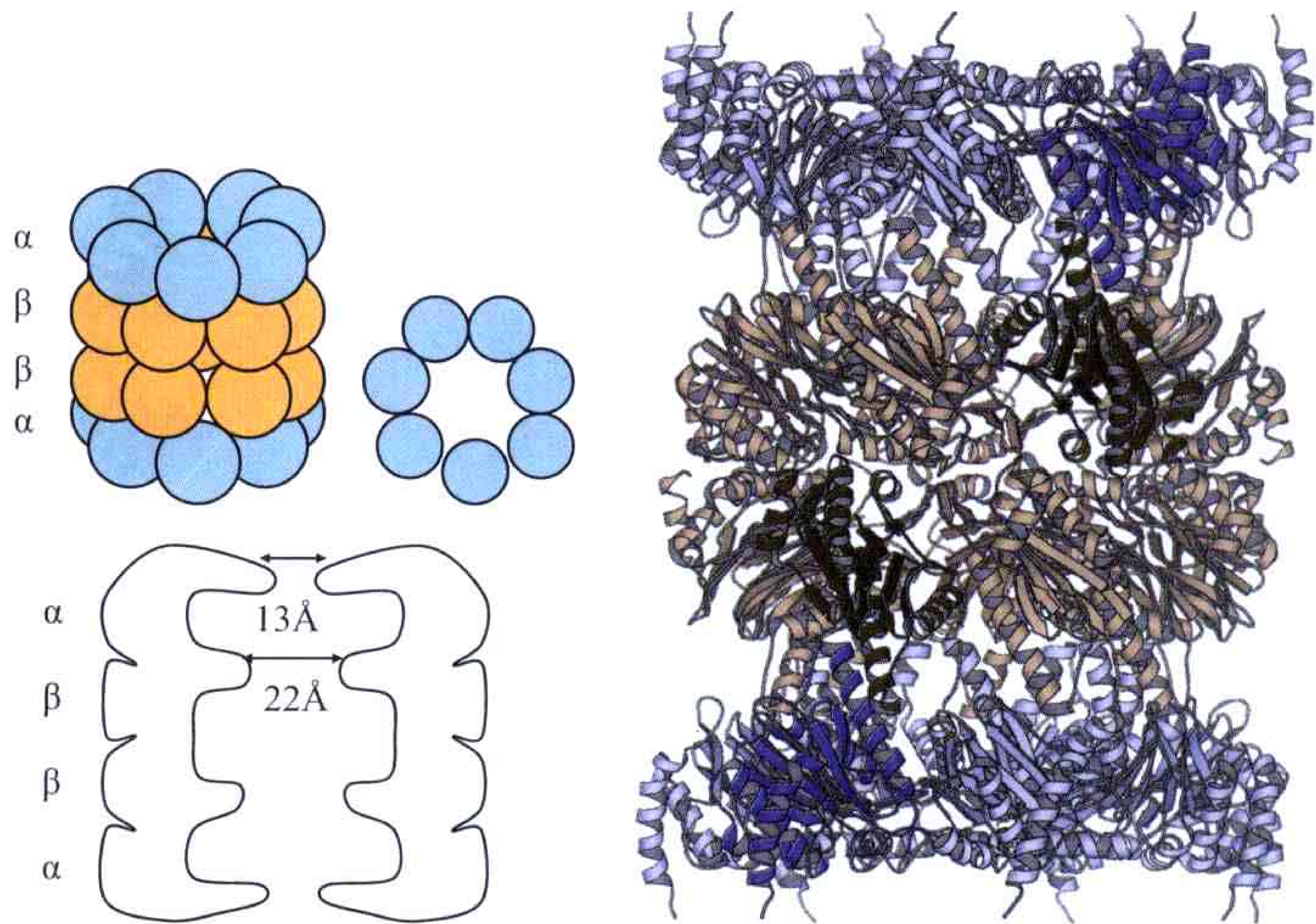


图 9.32　由 4 个环、每环有 7 个亚基组成的 20S 蛋白酶体结构。两个 β 环（棕色）位于中心，α 亚基（蓝色）构成外侧的环。左：进入小室的狭窄通道和随后进入中心酶解空腔的通道的示意图。活性位点由 β 亚基构成，面向酶解空腔。右：来自闪烁古生球菌的 20S 蛋白酶体的晶体结构。

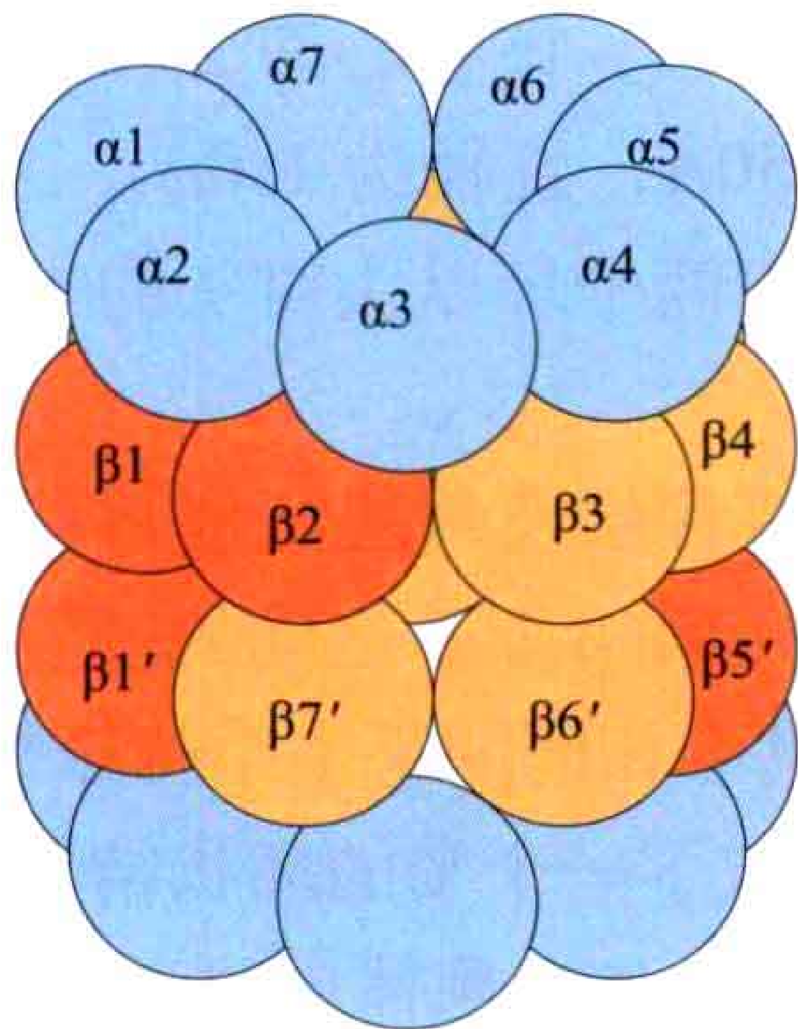


图 9.33　在真核生物 20S 蛋白酶体中，7 个 α 亚基（蓝色）和 7 个 β 亚基（棕色）都是独特的。该图显示了位于酶解空腔一侧的活性的 β 亚基（深棕色）1、2 和 1' 与位于相反一侧的亚基 5 和 5' 之间的空间关系。

所有细菌和绝大多数古细菌的蛋白酶体仅由一种类型的 α 亚基和一种类型的 β 亚基构成。对于一些古细菌和真细菌，有两种不同类型的 α 亚基和两种不同类型的 β 亚基。在真核生物中，α 亚基和 β 亚基均有具同源性的 7 种不同类型（图 9.33）。这对真核生物 20S 蛋白酶体的影响是：精确的七重对称被破坏。

在圆柱体的两侧，结合有调控性的 19S 复合体。在真核生物中，只有通过狭窄的通道及两端 19S 调控性复合体形成的通道入口才能进入被包围的内部空腔中。这些调控性复合体能阻止那些不受控制的蛋白质水解，并且负责识别、去泛素化、去折叠及将底物转运进空腔中进而降解。

蛋白酶体复合体能使天然折叠的泛素化蛋白变性，也能使那些非天然折叠状态的蛋白质变性。但是，蛋白酶体的 20S 部分只能降解那些去折叠了的蛋白质。α 亚基形成的环状结构直径是 13Å，这对于一个准备进入空腔的折叠状态的蛋白质来说太过狭窄了。因此，在底物被转运进酶解空腔之前，需要去折叠。这个功能是由 19S 复合体行使的，可以说是类似于分子伴侣的活性。19S 复合体可以解离为两个亚基，称为“基座”和“盖子”。“基座”包含 6 个 ATP 酶亚基，可能参与去折叠过程。它们属于 AAA+ 酶超家族（见 5.3 节）。“盖子”可能参与识别多聚泛素链。26S 蛋白酶体的 S5A/Pus1/RPN1 亚基（不同物种中有不同名称）可能定位于“基座”和“盖子”的互作面上，并通过一个小的泛素互作模体结合多聚泛素链。

在哺乳动物中，有一种称为免疫蛋白酶体（immuno-proteasome）的蛋白酶体。它参与产生那些被 I 类 MHC 分子呈递的特定的抗原肽。这个功能专一化了的蛋白酶体有一个结合于 20S 核心上的 11S 复合体。它被称为 PA26 或 11S 激活复合体。这个类型的蛋白酶体涉及抗原的处理过程。酵母的 20S 和 PA26 的复合体结构已得到研究（图 9.34）。在结合 PA26 之前，α 亚基的 N 端通过环状结构阻止底物进入空腔。七重对称的 PA26 使 7 个 α 亚基的 N 端发生重排。进而，环状结构打开开口，简化了底物进入酶解空腔的过程。

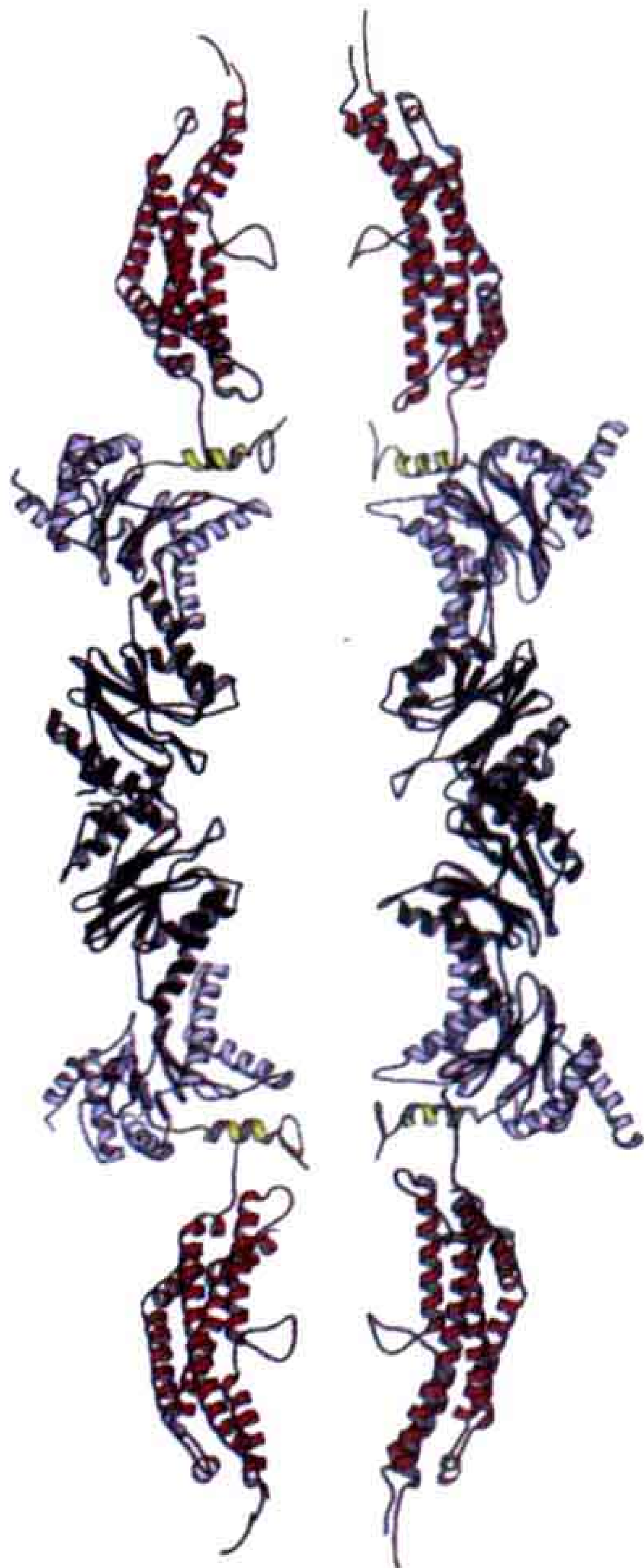


图 9.34　酵母 20S 蛋白酶体与 PA26 复合体结构的中心部分。PA26 即 11S 激活体（红色）。黄色部分是蛋白酶体 α 亚基 N 端发生重排的部分（PDB: 1Z7Q）。

9.2.3.1　活性位点

20S 蛋白酶体内部有三个空腔，其直径大约为 50Å。活性位点位于面对面排列的 β 亚基构成的中心空腔内。这个内侧空腔的入口直径大约为 30Å，而外侧空腔的入口更为狭窄。与 19S 调控性复合体的相互作用可能是使这些外侧空腔的环状结构打开的原因。

蛋白酶体的 β 亚基属于 Ntn- 水解酶（N-terminal nucleophile hydrolase）家族。它们总是以非活性的前体形式产生，然后通过内部的自催化剪切机制转化为活性形式。这个过程使苏氨酸（Thr1）暴露，使其成为 N 端氨基酸。水解过程中，苏氨酸行使亲核基团的功能，位于 N 端的游离氨基行使广义碱的功能。这个游离的 N 端氨基是所有 Ntn- 水解酶共同的特征。催化时苏氨酸的羟基氧与底物的 S1 残基的羰基碳共价连接。该共价键在随后的步骤中被水解。

被切下的前体肽有不同的长度和序列，但苏氨酸的前一位残基总是甘氨酸。对于自催化剪切过程，没有质子受体可供利用，因此需要一个水分子。苏氨酸附近的一系列保守氨基酸参与酶内部的相互作用，也参与与底物的相互作用。活性位点之间的距离大约为 28Å，这可能解释了为什么水解产生的肽段有如此长度。

真核生物中存在 7 种不同的 β 亚基，其中只有 β1、β2 和 β5 是有活性的酶。因此，与细菌中 2×7 个亚基均有活性相比，真核生物中只有 2×3 个亚基有活性。这些亚基能结合醛类抑制剂。这些亚基中一个 Thr1 的突变将会导致至少一个经典的蛋白酶体活性的丧失。β1 倾向于在谷氨酸后切割；β2 类似于胰蛋白酶，倾向于在精氨酸或赖氨酸后切割；β5 类似于糜蛋白酶，倾向于在芳香性氨基酸后切割。然而，这种特异性并不严格，并且真核生物的蛋白酶体可以切割绝大多数的肽键。两个 β 环的活性位点和非活性残基在结合底物及引导可能的切割上均表现出复杂的协同性。

蛋白酶体在不同的通路中对关键蛋白行使受调控的降解功能。因此，蛋白酶体的抑制剂受到广泛关注。细胞周期可以在各个阶段被捕获，这导致了细胞增殖的抑制。研究人员已经对许多抑制剂的结构和功能进行了分析。

9.2.3.2　寡聚化的受 ATP 调控的肽酶

许多大的寡聚化肽酶已被发现，它们是胞内酶。它们寡聚化的特性使它们倾向于降解肽类，而不是降解绝大部分的蛋白质。在一些情况下，这些大的肽酶把蛋白酶体产生的肽类降解为更短的肽类或氨基酸。许多寡聚化肽酶具有 ATP 酶活性，如 FtsH、Lon、ClpAP 及 ClpXP。ATP 酶或 ATP 酶结构域参与待降解蛋白质的去折叠。ATP 酶结构域之间是同源的，属于 AAA+ 超家族，并在结构上形成六元环。AAA+ 蛋白是化学机械性（chemo-mechanical）酶，它能改变其他蛋白质的构象，使大分子复合体解聚，或使蛋白质在细胞中移位。但是，蛋白质水解亚基或结构域可以不同。

AAA+ 组分由两个结构域组成：α 结构域和更大的 α/β 结构域。α 结构域的第一个螺旋是一个关键的元件。它可以伸直，以不同程度弯曲，或在中间形成两个残基的突起。

这将对 α 结构域和 α/β 结构域之间的空间与角度产生影响。α 结构域和 α/β 结构域之间的关系有赖于腺苷酸的状态。一个亚基的 α 结构域与邻近的另一个亚基的 α/β 结构域相互作用。这种相互作用根据核苷酸状态的改变而改变，并且这可能使底物肽或底物蛋白被拉入。AAA+ 组分六聚体的性质表明，就像 ATP 合酶一样，AAA+ 组分的亚基分别经历了 ATP 水解和 ADP → ATP 的交换过程，并以此来使底物蛋白变性和移位(见5.3节)。

表 9.3　寡聚化的依赖 ATP 的蛋白酶

蛋白质	蛋白质环的数目	AAA+ ATP 酶的寡聚化结构	蛋白酶的寡聚化结构	酶解空腔入口的直径 /Å	活性位点
LonA/B	2	6（7?）个结构域	6 个结构域	18	Ser+Lys
FtsH	2	6 个结构域	6 个结构域	20	Zn
ClpAP/ ClpXP	4	蛋白酶的桶状结构两侧各 6 个 ClpA 或 ClpX 亚基	2×7 个 ClpP 亚基	10	Ser, His, Asp
HslVU	4	蛋白酶的桶状结构两侧每侧各 6 个 HslU（或 ClpY）亚基	6×2 个 HslV（或 ClpQ）亚基	20	Thr, N 端 NH_2 基团
26S 蛋白酶体	6	6 个亚基	7α+7β+7β+7α 个亚基	13	Thr, N 端 NH_2 基团

9.2.3.3　HslVU

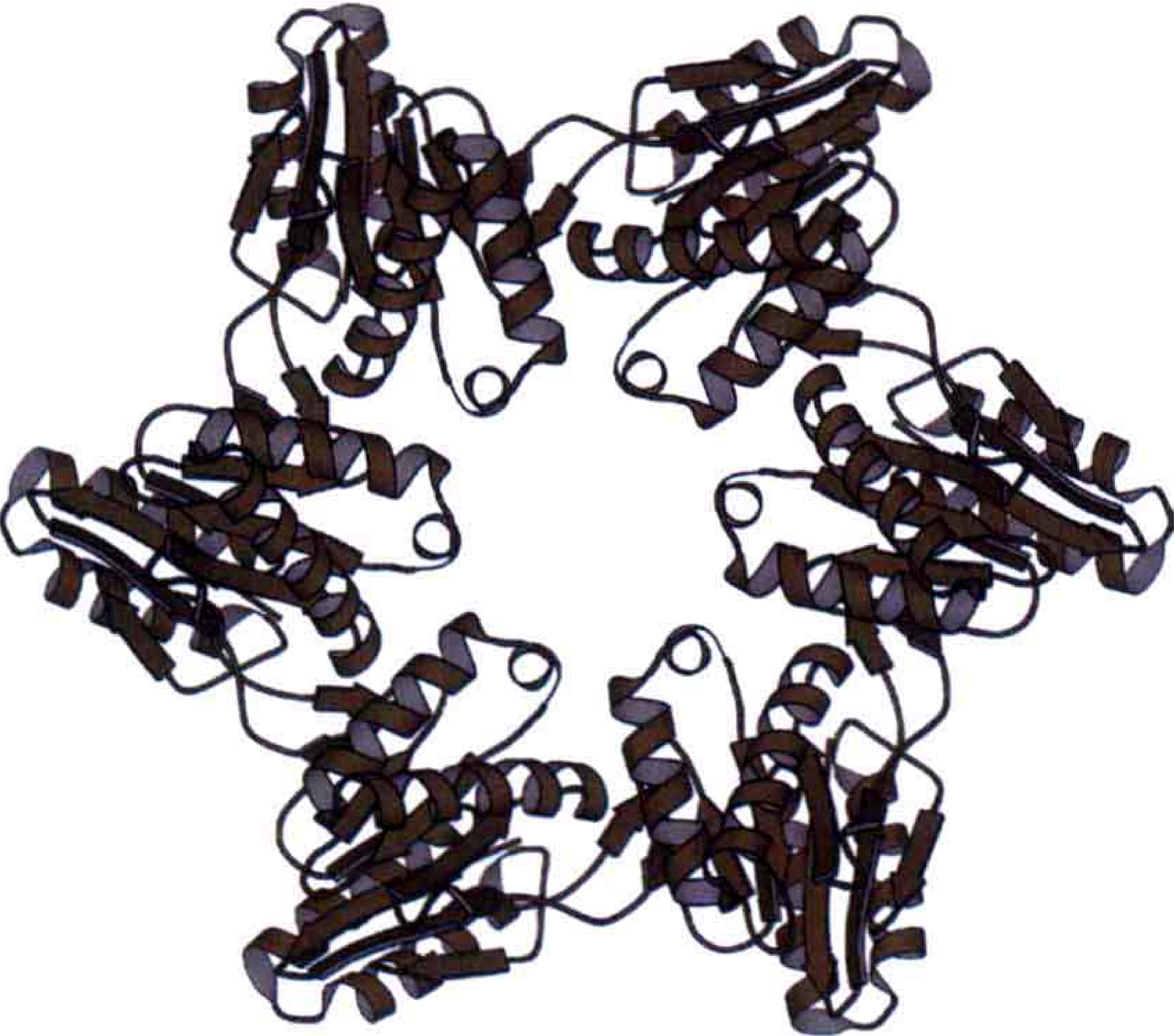


图 9.35　海栖热袍菌（*Thermotoga maritima*）HslV 的结构。它是 20S 蛋白酶体在细菌中的亲源蛋白。它由两个相同亚基组成的六元环构成（图中只显示了一个六元环）。这个蛋白质是一个苏氨酸蛋白酶，其亚基的结构与蛋白酶体 β 亚基的结构类似（PDB: 1M4Y）。

在一些细菌中，存在一组蛋白酶体的远亲。这组蛋白酶被称为 HslV（heat shock locus V）或 ClpQ。HslV 可以被热诱导，并可以降解非天然折叠的蛋白质。与 20S 蛋白酶体类似，蛋白酶颗粒由两个环（在这个例子中是同六聚体）组成，形成了一个用于水解蛋白质的空腔（图 9.35）。HslV 与蛋白酶体的 β 亚基具有一定的序列相似性（20%），但其 α 亚基是缺失的。这表明蛋白酶体的 β 亚基是更早期的形式，α 亚基是从这种早期的形式演化来的。HslV 是一种 Ntn- 水解酶，与蛋白酶体类似，它是以苏氨酸作为亲核基团的。它也在空腔内侧具有活性位点。对于短肽底物，HslV 自身具有很低的蛋白酶活性；而对于蛋白质，HslV 自身的活性可以忽略。

与 20S 蛋白酶体类似，HslV 与一个被称为 HslU 或 ClpY 的 AAA+ 超家族的 ATP 酶复合体相互作用。HslU 在结构上与蛋白酶体的 ATP 酶和 Hsp100 有关。HslU 可使 HslV 的活性增强 1~2 个数量级。HslU 也由六元环组成，结合在 HslV 双环的两侧（图

9.36）。HslU 是分子伴侣 Clp/Hsp100 家族的一员。这些蛋白质既参与蛋白质聚集体的解聚，也参与靶蛋白的降解。这两个看起来相反的活性之间共同的功能是使蛋白质去折叠，一种情况是使蛋白质降解，另一种情况是使蛋白质复性。在这两种情况下，均需要消耗 ATP。

HslU 单体有三个结构域：N 端结构域，（参与结合 ATP）、中间结构域（I）及 C 端结构域。ATP 结合结构域面向外侧并远离 HslV 亚基。HslU 与 HslV 之间的相互作用可诱导构象变化，这可能是使复合体活性增加的原因。HslV 双环与 HslU 单环的复合体结构显示出一种结构排列的不对称性。在结合 HslU 环后，靶蛋白移位时所通过的环状结构变得非常宽大，其直径几乎达到了 20Å。两个 HslV 环之间的互作界面是疏水性的，而 HslV-HslU 之间的互作界面极其亲水，具有很多氢键。HslU 的 N 和 C 结构域可能负责复合体的装配和激活，而 I 结构域可能参与选择底物并引导底物进入 HslV 内部的酶解空腔中。HslV 的肽酶活性也需要 HslU 才能被激活，但激活机制的细节并不被完全了解。

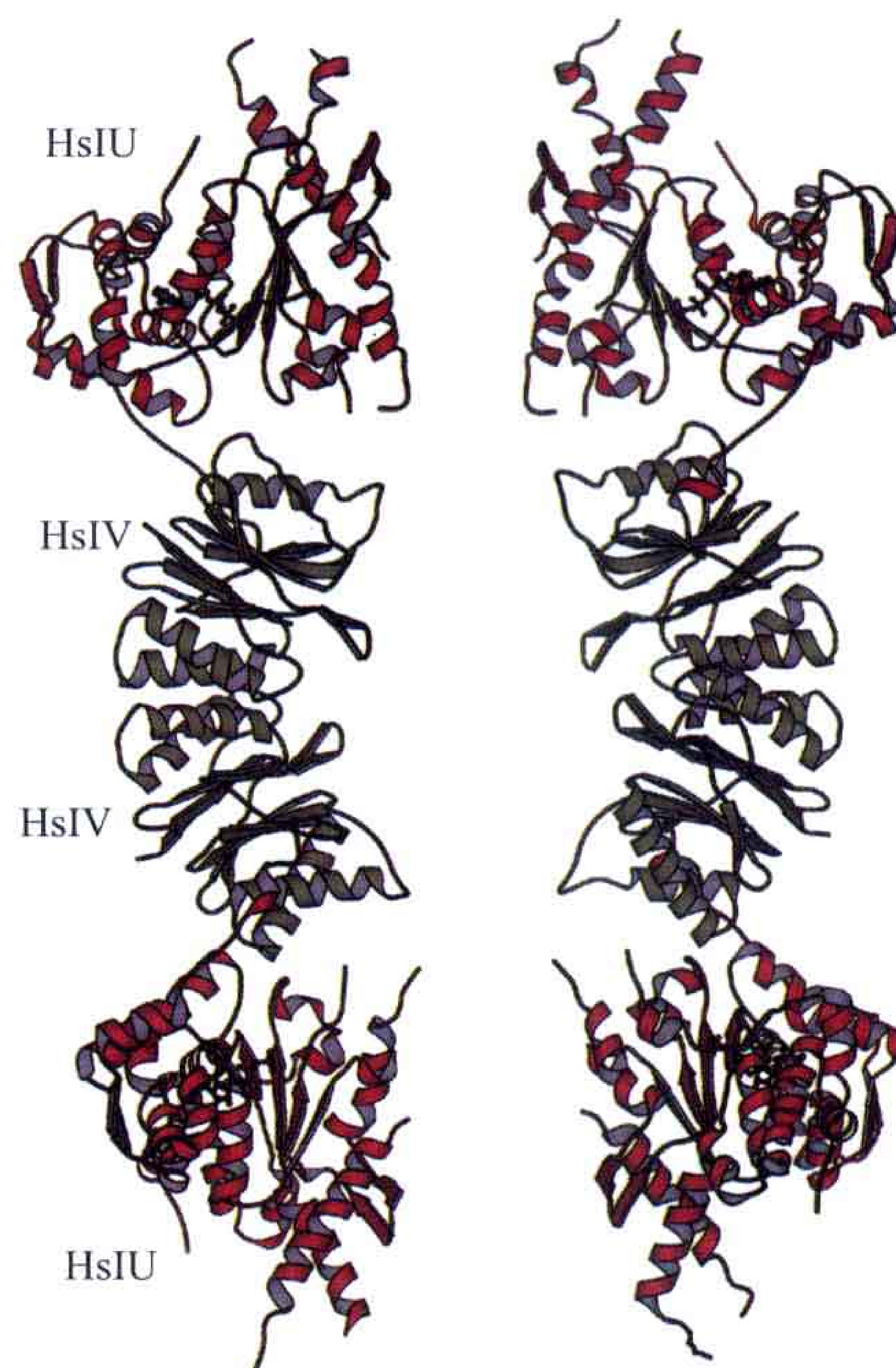


图 9.36　来自大肠杆菌的 HslV 和 HslU 复合体。图中只显示了一对亚基。具有蛋白质水解活性的 HslV 亚基两侧可以与一个或两个具有 ATP 酶活性的 HslU 环互作。如果只有一个 HslU 环与复合体互作，那么与 HslU 互作的 HslV 环称为顺式环，而远端的 HslV 环称为反式环。完整的该复合体是完整的蛋白酶体的亲缘蛋白。

9.2.3.4　FtsH

FtsH 与其他 AAA+ 蛋白酶有以下几个方面的不同：它结合在细胞质膜的内侧；在细菌中它是必需且保守的；它在线粒体和叶绿体中发挥重要作用；它在蛋白质的质量控制和膜蛋白的降解中起重要作用。ATP 的能量用于将待降解的蛋白质拉出细胞膜并使它们去折叠。跨膜区，AAA+ 结构域，以及蛋白酶结构域包含在同一条肽链内。海栖热袍菌 FtsH 的结构排列为两个环形。

蛋白酶环是一个平面的六边形，其结构域均由螺旋构成（图 9.37）。AAA+ 结构域形成一个环状。与蛋白酶环形成的几乎完美的六边形相比，AAA+ 环具有更大的柔性且并不完美。其对称性更恰当地说是二重对称的。活性位点包裹在两环之间，包含一个与两个组氨酸及一个天冬氨酸残基配位的锌离子。组氨酸经常出现在蛋白酶或“zincin”的锌结合模体中，即 HEXXH 模体。活性位点之间的距离约为 34Å。AAA+ 结构域之间裂隙的宽度为 20Å。在 AAA+ 环结合和水解 ATP 及把目的蛋白拉入并去折叠这一功能性循环过程中，AAA+ 环很可能经历了对称性和构象的改变。

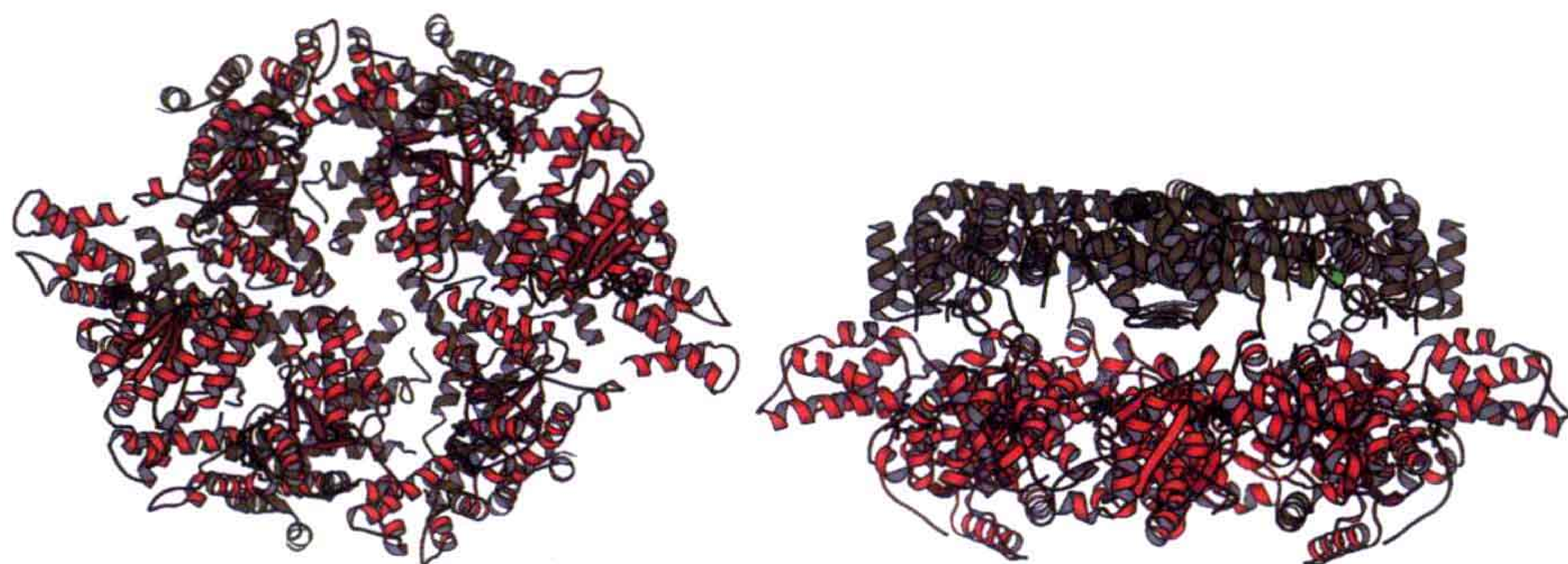


图 9.37　来自海栖热袍菌的 FtsH 结构。AAA+ 结构域位于上面的 FtsH 的俯视图（左侧）及侧视图（右侧）。FtsH 的两个环由 6 条多肽链构成。蛋白酶结构域（棕色）位于上面，AAA+ 结构域（红色）位于下面。活性位点位于内侧空腔内。AAA+ 结构域通过两个跨膜螺旋与细胞膜相连接（PDB: 2CE7）。

蛋白酶数据库

http://merops.sanger.ac.uk

延伸阅读（9.1 节）

原始文献

Ali MMU, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**: 1013-1017.

Baram D, Pyetan E, Sittner A, Auerbach-Nevo T, Bashan A, Yonath A. (2005) Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proc Natl Acad Sci USA* **102**: 12017-12022.

Guo F, Maurizi MR, Esser L, Xia D. (2002) Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *J Biol Chem* **277**: 46743-46752.

Howard BR, Vajdos FF, Li S, Sundquist WI, Hill CP. (2003) Structural insight into the catalytic mechanism of cyclophilin A. *Nat Struct Biol* **10**: 475-481.

Jiang J, Prasad K, Lafer EM, Sousa R. (2005) Structural basis of interdomain communication in the Hsc70 chaperone. *Mol Cell* **20**: 513-524.

Kennaway CK, Benesch JLP, Gohlke U, Wang L, Robinson CV, Orlova EV, Saibil HR, Keep NH. (2005) Dodecameric structure of the small heat shock protein Acr1 from *Mycobacterium tuberculosis*. *J Biol Chem* **280**: 33419-33425.

Lakshmipathy SK, Tomic S, Kaiser CM, Chang H-C, Genevaux P, Georgopoulos C, Barral JM, Johnson AE, Hartl FU, Etchells SA. (2007) Identification of nascent chain interaction sites on trigger factor. *J Biol Chem* **282**: 12186-12193.

Li J, Xinguo Q, Sha B. (2003) The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate. *Structure* **11**: 1475-1483.

Stamler R, Kappé G, Boelens WM, Slingsby C. (2005) Wrapping the α-crystallin domain fold in a chaperone assembly. *J Mol Biol* **353**: 68-79.

Xia D, Esser L, Singh SK, Guo F, Maurizi MR. (2004) Crystallographic investigations of peptide binding sites in the N-domain of the ClpA chaperone. *J Struct Biol* **146**: 166-179.

Xu Z, Horwich AL, Sigler PB. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388**: 741-750.

综述文章

Gruber CW, Cemasar M, Hreas B, Martin JL, Craik DJ. (2006) Protein disulfide isomerase: The structure of oxidative folding. *TIBS* **31**: 455-464.

Harrison CJ. (1997) La cage aux fold: Asymmetry in the crystal structure of GroEL-GroES-(ADP)7. *Structure* **5**: 1261-1264.

Haslbeck M, Franzmann T, Weinfurtner D, Buchner J. (2005) Some like it hot: The structure and function of small heat shock proteins. *Nat Struct Mol Biol* **12**: 842-846.

Hendrickson WA, Lin Q. (2008) Exchange we can believe in. *Structure* **16**: 1153-1155.

Lin Z, Rye HS. (2006) GroEL-mediated protein folding: Making the impossible, possible. *Crit Rev Biochem Mol Biol* **41**: 211-239.

Saibil HR, Ranson NA. (2002) The chaperonin folding machine. *TIBS* **27**: 627-632.

Stirling PC, Bakhoum SF, Feigl AB, Leroux MR. (2006) Convergent evolution of clamp-like binding sites in diverse chaperones. *Nat Struct Mol Biol* **13**: 865-870.

Young JC, Agashe VR, Siegers K, Hartl FU. (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nature Rev* **5**: 781-791.

延伸阅读（9.2 节）

原始文献

Föster A, Whitby FG, Hill CP. (2003) The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. *EMBO J* **22**: 4356-4364.

Löwe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R. (1995) Crystal structure of

the 20S proteasome from the archeon *T. acidophilium* at 3.4Å resolution. *Science* **268**: 533-539.

Nickell S, Mihalache O, Beck F, Hegerl R, Korinek A, Baumeister W. (2007) Structural analysis of the 26S proteasome by cryoelectron tomography. *Biochem Biophys Res Commun* **353**: 115-120.

综述文章

Baker TA, Sauer RT. (2006) ATP-dependent proteases of bacteria: Recognition logic and operating principles. *TIBS* **31**: 647-653.

Ciechanover A. (2004) Intracellular protein degradation: From a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. In: T Frängsmyr (ed.), *Les Prix Nobel*, The Nobel Foundation, pp. 1197-1211.

Groll M, Bochtler M, Brandstetter H, Clausen T, Huber R. (2005) Molecular machines for protein degradation. *Chem Biochem* **6**: 222-256.

Groll M, Huber R. (2004) Inhibitors of the eukaryotic 20S proteasome core particle: A structural approach. *Biochem Biophys Acta* **1695**: 33-34.

Heinemeyer W, Ramos PC, Dohmen RJ. (2004) The ultimate nanoscale mincer: Assembly, structure and active sites of the 20S proteasome core. *Cell Mol Life Sci* **61**: 1562-1578.

Huntington JA. (2006) Shape-shifting serpins — Advantages of a mobile mechanism. *TIBS* **31**: 427-435.

Reinstein E, Ciechanover A. (2006) Narrative review — Protein degradation and human diseases: The ubiquitin connection. *Ann Intern Med* **145**: 676-684.

Robinson PA, Ardley HC. (2004) Ubiquitin-protein ligases. *J Cell Sci* **117**: 5191-5194.

Schwartz DC, Hochstrasser M. (2003) A superfamily of protein tags: Ubiquitin, SUMO and related modifiers. *TIBS* **28**: 321-328.

Whisstock JC, Bottomley SP. (2006) Molecular gymnastics: Serpin structure, folding and misfolding. *Curr Op Struct Biol* **16**: 761-768.

（王子曦 译　廖美雪　苏晓东 校）

第十章
膜蛋白

10.1 介绍

生物膜不仅把细胞和周围环境分离开，也把细胞内各个腔室彼此分隔开。如第四章所述，生物膜通常被描述为流动镶嵌模型，这是由 Singer 和 Nicolson 在 1972 年提出的，认为膜蛋白镶嵌在磷脂组成的脂双层里面。作为一道水溶性物质和电荷的屏障，生物膜的存在让细胞能够把细胞组分和资源保存在一个独立的、可控制的腔室里面，从而将细胞内保持在一个高度受控的物理化学环境，同时也维持了跨膜的电化学梯度和电势梯度。这些跨膜梯度是信号转导和次级转运的基础，对细胞和周围环境的相互作用至关重要。

膜蛋白具有和水溶性蛋白相似的功能，可以作为酶、载体或者结构单位等。其存在使生物膜具有生化功能和可调控性。因此膜蛋白具有重要的生理功能，对于生命过程非常重要。膜蛋白同时具有自身独特的功能，如可以作为转运体、通道、受体或者结构单元。转运体可做主动转运，即通过消耗某种能量，逆浓度地定向转运物质。主动转运造成并维持着跨膜的渗透压和电化学梯度。当利用的能量是外源的，如来自 ATP 水解，或者光反应或氧化还原反应，这类蛋白质就称为初级转运体。如果利用的能量是已经建立好的跨膜梯度（通常是钠离子或者质子梯度），那么这类蛋白质就叫次级转运体。次级转运体又可进一步根据被转运底物跨膜运输的协同关系分为共转运体或者反转运体（后者又常被称为交换体）。通道蛋白（包括膜孔）能够协助物质从高浓度到低浓度的被动运输，并且能和调控跨膜通道开关的门控机制偶联起来。研究比较透彻的通道蛋白有离子特异性通道、膜孔蛋白和机械力敏感通道。受体蛋白以不同的状态调控跨膜信号转导，并且利用与特定底物（如激素、神经递质或者气味因子）的结合来稳定信号传递的结构（见第十一章）。一些受体同时也是门控通道，如亲离子型谷氨酸受体和电压门控钾离子通道。

膜蛋白也是病理学中主要的药物靶标。在细胞内，膜组分通常只占细胞体积的一小部分，而且膜蛋白相比于水溶性蛋白表达量通常较低。同时，膜蛋白通常在膜外就不稳定，导致其难于被研究。由于这些或者其他的原因，膜蛋白组学和膜蛋白的结构研究还只是慢慢在发展。不过由于有了强大的同步辐射光源和完善的电脑晶体学软件，膜蛋白的结构研究发展得越来越快。

在图 10.1 中展示了不同种类的膜蛋白。外周膜蛋白只是和膜表面有相互作用。内

在膜蛋白则是跨膜蛋白或者蛋白质通过脂肪链或糖磷脂酰肌醇（GPI）锚定在膜上。本章将讨论跨膜蛋白的结构和性质。

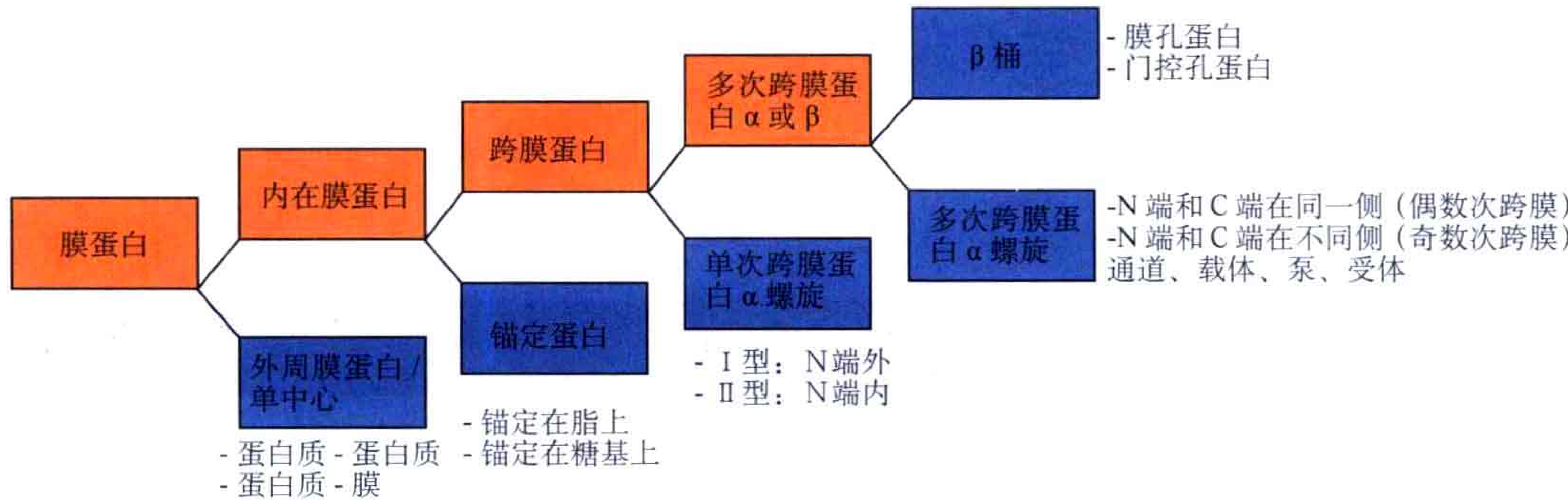


图 10.1　不同种类的膜蛋白。

10.1.1　从膜组分到内在膜蛋白

在 1948 年，Benjamin Libet 发现了一个结合在膜上的酶——一个来自大王乌贼神经细胞的 ATP 酶。在同一时期人们对于神经动作电位的理解也有了重大突破，认为其是由离子跨膜流动导致的。在 Hodgkin 和 Huxley 于 1952 年发表的一系列令人印象深刻的文章的基础上，离子通道存在的假说首次被提出来。Libet 发现的来自大王乌贼神经的酶在 1957 年被 Jens Christian Skou 证明可以被特定比例的钠离子和钾离子激活到最高活性，Skou 随后推断这个酶就是人们寻找多时的钠钾泵。钠钾泵存在的假说早在 1890 年就被 Ernest Overton 提出。不过，关于钠钾泵就是一个镶嵌在膜里面的酶的概念在 1957 年时还难以被理解——内在膜蛋白的概念只有到 20 世纪 60 年代时才发展起来。在 50 年代后期普遍的观点是由于疏水作用，蛋白质折叠只有在水溶液环境下才可能发生。这个观点是从第一个蛋白质结构，即肌红蛋白结构中发展出来的。

在 20 世纪 60 年代，Jardetzky 提出一个在概念上的如何完成主动转运的“交互接触模型”，即一个跨膜的系统，其高亲和力结合位点暴露向膜的某一面，而当其转换成另一个状态时，朝膜的另一面便暴露出低亲和力位点，这样就可有效地形成转运单元。此外，当这些状态是被其他过程如 ATP 水解来分开的时候，只要向这个系统提供定向组分，便可以保证一个单向的转运模式。现在很多例子证明了这个模型的正确性，如膜蛋白里面被研究得最为透彻的 P 型 ATP 酶泵。相似地，通道蛋白不久前被认为对不同离子有不同的通透性。与蛋白质泵不一样的是，通道蛋白并不利用高亲和力位点，而是为特定离子的快速通过提供一个选择体系，从而造成离子浓度和膜电位的快速变化。

10.1.2　第一个膜蛋白结构

第一个膜蛋白的精细结构来自 Richard Henderson 和 Nigel Unwin 在 1975 年对“紫色细胞膜”即细菌视紫膜质的二维晶体所做的影响深远的电镜研究。从一系列电子衍

射图里，他们得到了一个分辨率为 7Å 的三维密度图，在图里可见 7 个杆状密度，各有 35~40Å 长，横贯细胞膜（图 10.2）。根据观察到的密度体积结合蛋白质分子质量的大小，他们得出结论认为这些柱状物可以描述为跨膜 α 螺旋堆积在一起并构成膜蛋白的结构基础。

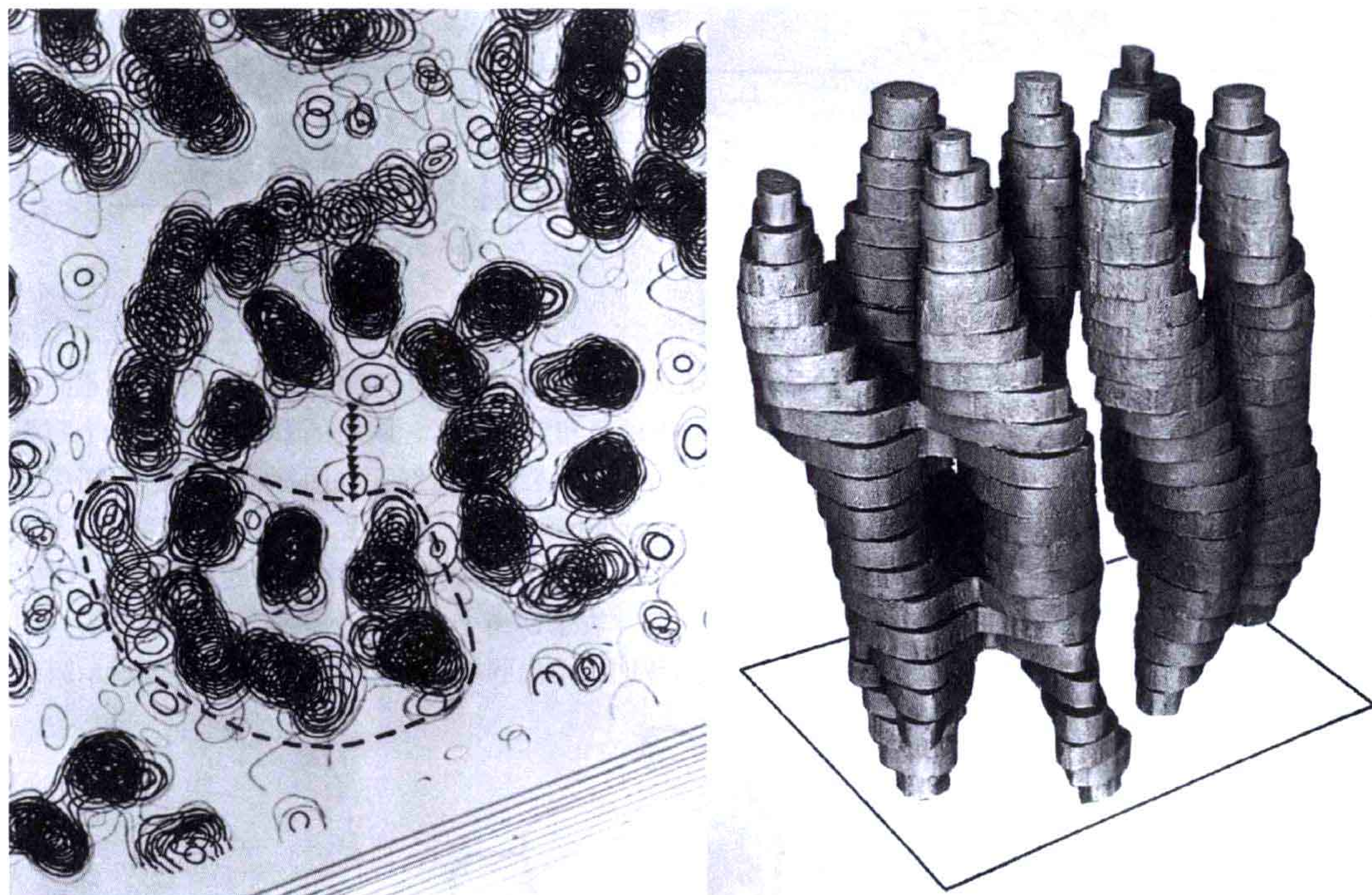


图 10.2　Henderson 和 Unwin 在 1975 年利用电镜对“紫色细胞膜”二维晶体研究得到的投影图（左）和从一系列倾斜样本中得到的细菌视紫红质的三维重构图（右）[原图来自从 Henderson R, Unwin PNT. (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257:**28-32. Copyright (1975) Nature. 在获得允许下的重印]。

第一个膜蛋白的三维晶体（细菌视紫红质）是在 1980 年由 Österhelt 和 Michel 得到的。在 1982 年，Michel 还第一次发表了对来自 *Rhodopseudomonas viridis* 的光合作用反应中心结晶的研究工作——这在不久之后被证明是非常明智的举动。尽管经历了好几次失败，Michel 在改变去垢剂和影响微团的添加剂之后获得了更好的晶体，进而成功地进行了 X 射线衍射。事实上当时大多数人认为他试图获得膜蛋白结构是不可能成功的。获得光合作用反应中心及其所有配体和金属催化中心的原子结构开辟了跨膜电子链和传递通道及其与脂分子紧密相互作用的新天地——即使以今天的标准看这也是一个巨大的成就。Michel 和他的同事 Deisenhofer 和 Huber 因此在 1988 年分享了诺贝尔奖（图 10.3）。早在 1980 年，Garavito 和 Rosenbusch 就报道了大肠杆菌外膜膜孔蛋白（后来被命名为 OmpF）的三维晶体能够在 X 射线下衍射到中等分辨率。不过直到 1992 年这个膜蛋白的结构才被解析出来。其结构显示 16 个 β 片层组成一个镶嵌在膜里面的大桶，这是跨膜孔道极具代表性的例子（图 10.4；另见第二章和附录 C）。

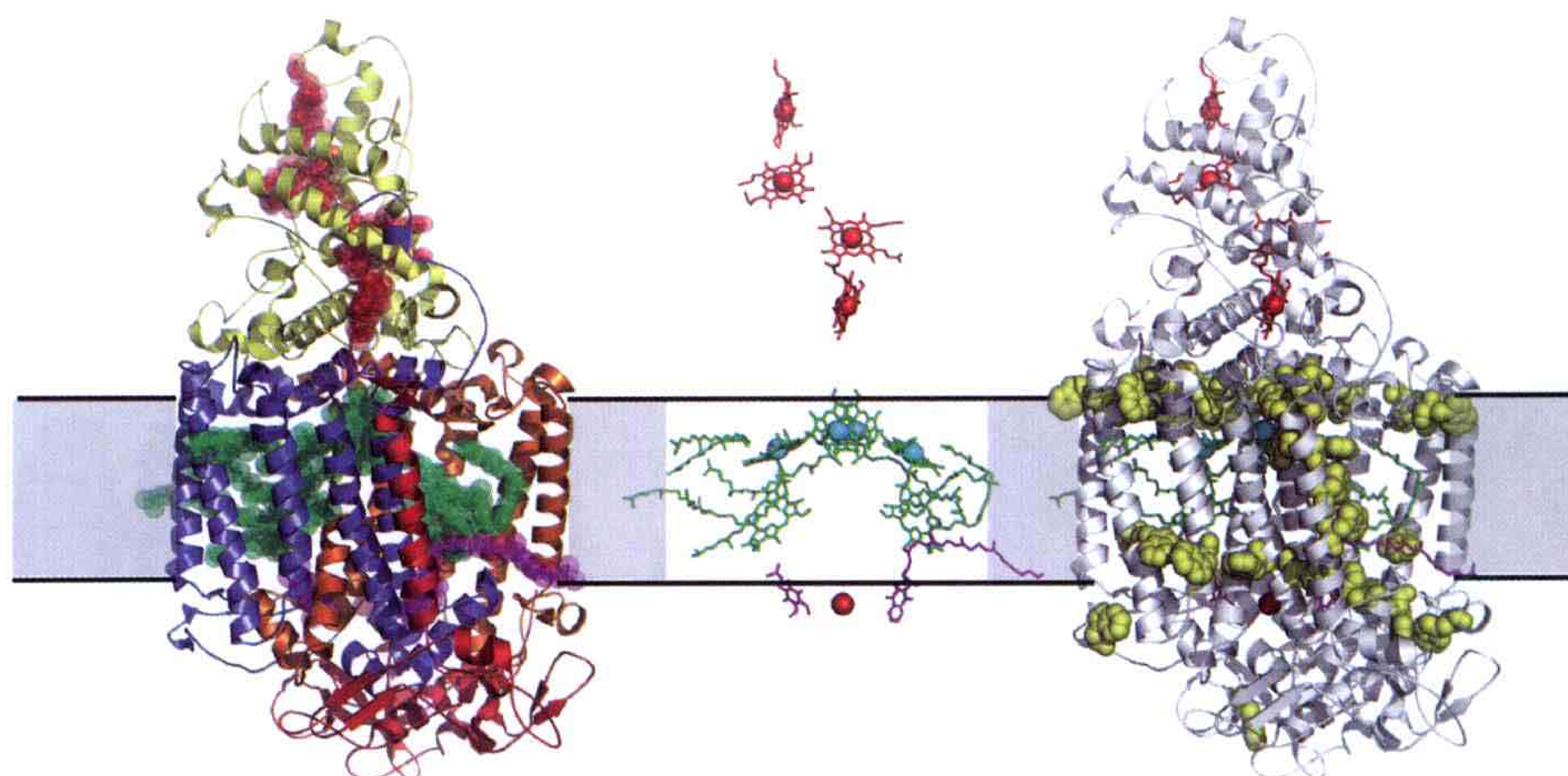


图 10.3　第一个膜蛋白复合体的原子结构——来自 *R. viridis* 的光合作用反应中心。左：卡通示意图。细胞色素亚基 C 用黄色表示，跨膜亚基 L 和 M 分别用橙色和蓝色表示，胞质区 H 亚基（有一个跨膜螺旋）用红色表示。电子传导配体用半透明球指示，其中 C 亚基的血红素基团标记为红色。细菌叶绿素（bacteriochlorophyll）和细菌脱镁叶绿素（bacteriopheophytin）标记为绿色，醌类标记为紫色（PDB 号：1PRC）。注意 L 和 M 亚基之间及从中心的叶绿素“特殊对”分成两支的配体间的赝对称性。中：同样的结构，但是只用棍状模型展示配体，用天蓝色小球表示镁离子，红色小球表示铁离子。右：同一个结构展示色氨酸（黄色）在跨膜亚基的分布，清楚地指示出膜界面的位置。灰色的方框表示膜的大概位置。

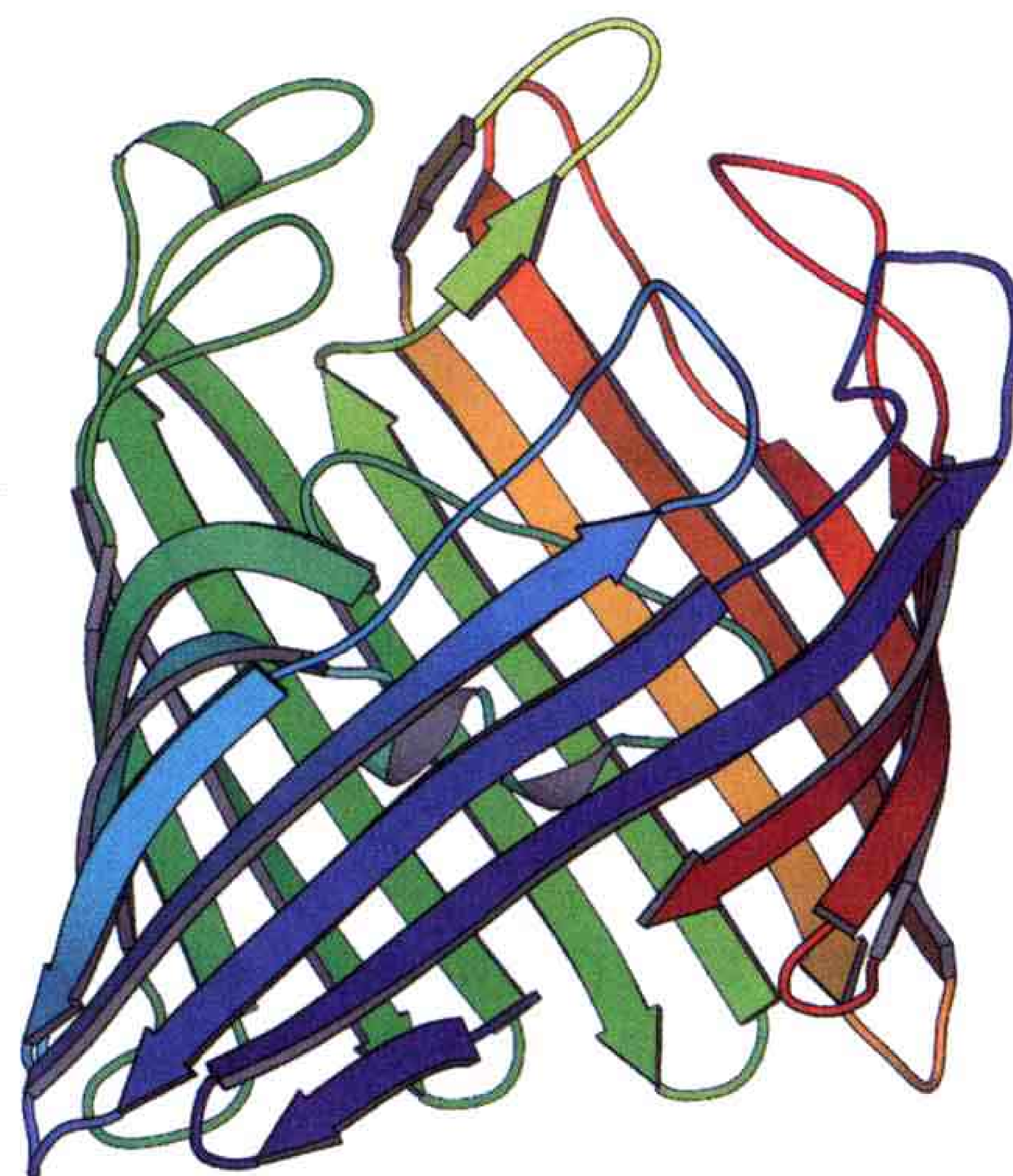


图 10.4　之后被命名为 OmpF 的膜孔蛋白的结构展示了一个跨膜的 β 桶型结构（PDB：2OMF）。

10.2　α 螺旋型和 β 桶型的膜蛋白结构

第一批的膜蛋白结构——细菌视紫红质（见 10.5.1 节）、光合作用反应中心（图 10.3）和膜孔蛋白 /OmpF（图 10.4），也说明存在两个本质上不同的膜蛋白类型：α 螺旋型和 β 桶型结构的蛋白质。这两类结构满足了一个基本的要求：减少蛋白质结构由

于定位在膜内油性环境而非水溶液环境造成的能量消耗。多肽链骨架在羧基和氨基的位置有形成氢键的趋势，这一点在跨膜区将会通过 α 螺旋或 β 桶以分子内形成氢键的方式系统性地得到满足。不过此时这两种结构又非常不一样：β 桶的特点是不同片层间的“长程”氢键，这样为保持氢键完整就没有多少空间可以进行构象变化。而 α 螺旋形成的是区域性的（n+4）氢键，这在跨膜螺旋存在的形式下允许进行构象变化（图 10.5）。确实，β 桶型的蛋白质大多数作为膜孔协助跨膜的被动运输，而更复杂的跨膜转运模式，如离子通道和转运蛋白，是基于 α 螺旋型结构的。

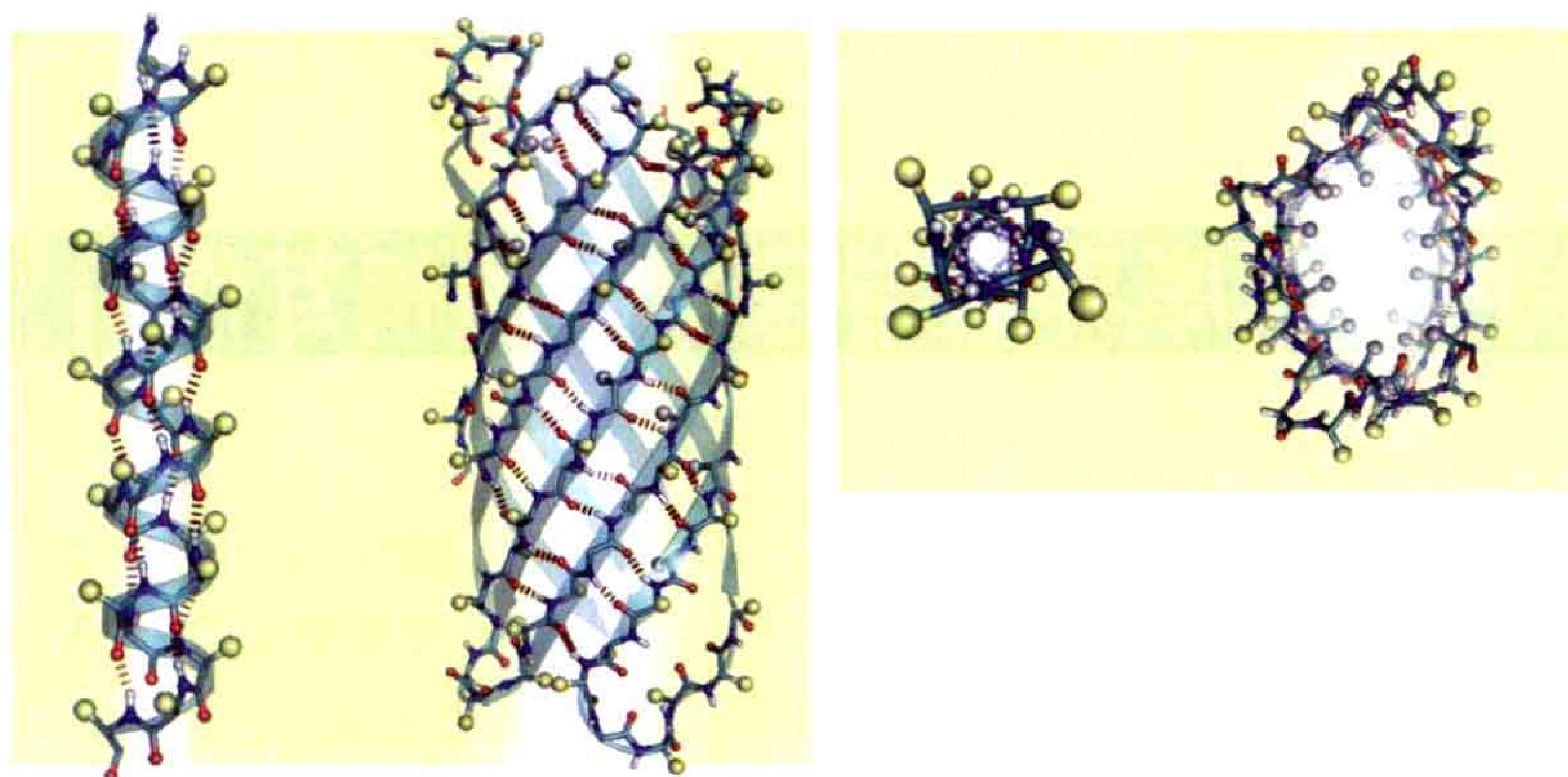


图 10.5　α 螺旋型和 β 桶型结构的氢键示意图比较（左）显示了完全不同的侧链（用黄色小球表示）向脂相暴露的模式（右）。图中 α 型和 β 型结构大小不是在同一个尺度。α 螺旋型结构展示的是一个有 21 个残基的跨膜螺旋。

α 螺旋型和 β 桶型跨膜片段的另一重要差异是它们的序列模式。α 螺旋型的蛋白质在整个跨膜区都需要(绝大多数)疏水侧链；而 β 桶型的则每隔一个残基才会接触到脂相，从而序列会被疏水性选择所限制。这对利用序列分析识别跨膜蛋白时有重要影响，目前已经发展出了非常好的序列分析软件来寻找跨膜螺旋，最显著的是有疏水特征的 21 个残基的片段——这个数值乘以在 α 螺旋中每个残基沿螺旋长轴上升的 1.5Å 就得到了脂双层的大概厚度，从而代表了跨膜螺旋的最小长度（见附录 C）。

10.2.1　单跨膜 α 螺旋膜蛋白和 monotopic 膜蛋白

α 螺旋形成一个独立的折叠单元。如果它具有疏水侧链并在两端有极性侧链，它就形成了一个稳定的插入膜的片段，极性的侧链作为在膜内侧和外侧界面的浮标，而疏水的侧链插入脂双层内部的脂相。与此同时，插入膜的片段，当其多肽链暴露于脂肪链前，没有氢键受体或者供体时才更可能偏爱多肽螺旋结构。这段多肽链在水溶性环境中反而不容易形成螺旋结构。

相对于一个伸展的肽段，α 螺旋能形成蛋白质的一个清楚的独立跨膜模块。把蛋白质锚定于膜上，而 N 端和 C 端位于膜的某一侧。只有一个穿过膜的 α 螺旋的蛋白质称为单次跨膜蛋白。1 型的单次跨膜膜蛋白其 N 端在胞外（最为普遍），而 2 型则其 N

端在胞内。

单次跨膜常常通过其跨膜螺旋形成二聚体甚至寡聚体。不幸的是，关于单次跨膜膜受体的跨膜结构域相互作用的结构信息还非常少（相比于它们的胞内和胞外结构域），目前没有完整复合物的晶体结构。

许多受体及与膜相连的酶是内在 monotopic（单中心）蛋白。这个词指蛋白质（或者结构域）不穿膜而是与膜一侧相连。它们需要去垢剂来溶取和结晶。例如，前列腺素 H2 合成酶 1 和 2，或者肉毒碱 O 型棕榈转移酶。一些膜蛋白同时具有一个或者几个跨膜螺旋和 monotopic 结构域，这种类型的离子有单胺氧化酶和脂肪酸氨基水解酶。

10.3 脂质及水脂界面与膜蛋白的相互作用

对于一些膜蛋白的早期研究表明，脂双层不仅是一个惰性的介质或者一个简单的屏障，脂质的属性，如链长、非饱和键和头部造成的整体性质影响着膜蛋白活性。钙泵活性和脂质性质的关系曾被仔细研究，脂质性质涉及脂肪链的长度（如 C20、C18、C16 和 C14）、脂肪链的结构（如饱和的或者单不饱和的）及头部基团的结构（如卵磷脂或者磷脂酰丝氨酸）。研究表明酶的活性与这些因素紧密相关。这种脂双层成分和膜蛋白嵌入及活性间的交互依赖性可能代表控制生物膜活性和膜蛋白分布的重要机制，也有可能是一些现象如脂筏的潜在推动力。

有的膜蛋白晶体里面脂分子（或者去垢剂）在蛋白质跨膜表面形成一个有序的整体。这在视紫红质的高分辨晶体结构中就被观察到过（图 10.6）。这类膜蛋白晶体形式特别有利于加深对蛋白质 - 脂相互作用的理解。

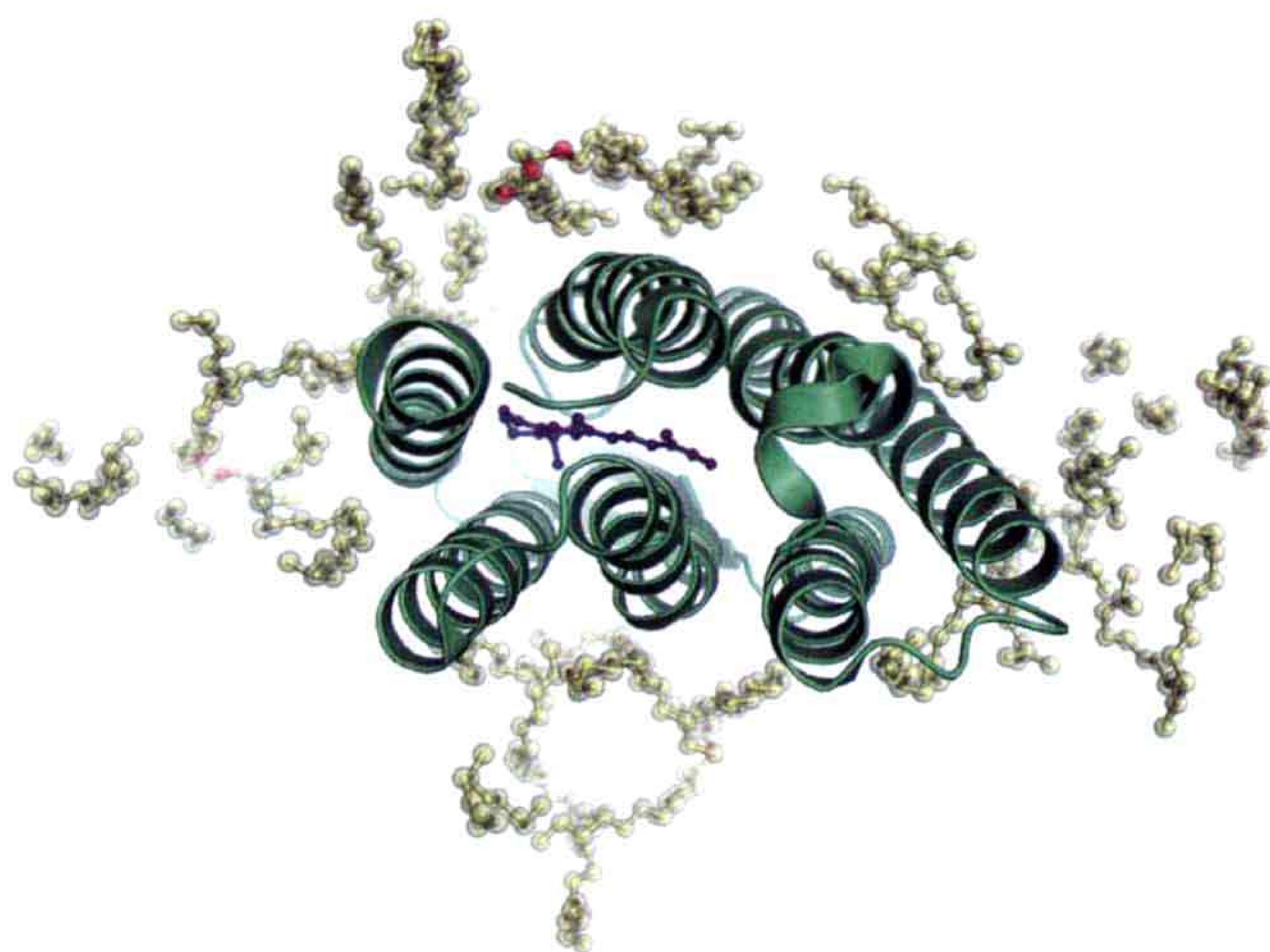


图 10.6　视紫红质结构周围的脂分子。从这个结构几乎可以完整地看到膜蛋白的脂化状态，这是理解蛋白质 - 脂质 - 水界面的复杂性质的基础。视网醛分子用蓝色标识，脂分子用黄色（碳原子）和红色（氧原子）标识（PDB：1C3W）。

10.4　膜蛋白催化的跨膜转运及信号转导

生物膜一个基本功能是它们具有通过膜转运蛋白的活动控制跨膜运输的能力。跨膜转运可能是通过膜孔和通道的被动运输，也可能是通过泵和转运蛋白的逆电化学梯度的主动运输（需消耗能量）（图 10.7）。细胞转运系统的基本特征包括运输的特异性和可调节性。不过蛋白质通道涉及的扩散速率和主动转运蛋白涉及的构象变化在时间尺度上有巨大差异。主动转运蛋白在任何情况下都不会允许一个同时向膜两侧开放的通道存在，因为这会轻易地平衡了梯度，让花在转运上的工作全无效果——就像一个打气筒在活塞处漏气一样。主动转运蛋白是通过大的构象变化来控制的，这些变化造成向内和向外的构象，由中间态或者过渡态的封闭状态将之隔开。其结果就是转运蛋白的转换速率很慢，在毫秒量级（或者更慢）。而另外，当门控机制打开了通道之后，通道蛋白就能提供一个完整的跨膜通路，其转运速度仅由扩散速率来限制。其转换速率在微秒和皮秒量级，这使快速的传送和反应成为可能。这方面的一个例子是电压门控的通道参与了神经组织里动作电位的形成和传导。

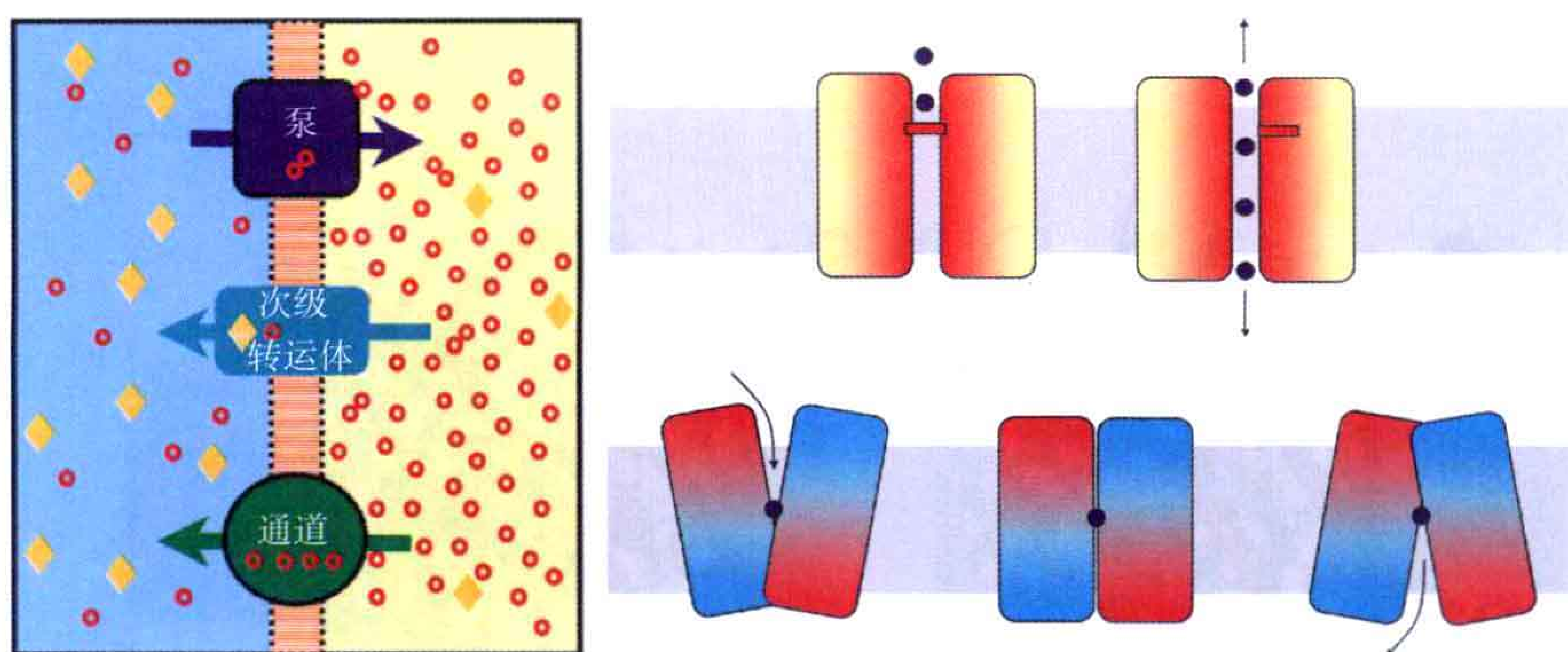


图 10.7　左：转运机制概述示意图。上：初级主动运输（一个泵）构建了红色阳离子的电化学梯度；中：一个次级转运体利用这种电化学梯度来主动地共转运黄色的溶质（如其他离子、代谢产物、糖和神经介质）；下：一个通道允许红色阳离子的顺势转运，其速率由离子通过选择性滤器时的扩散速率限制。右：门控通道和主动转运蛋白的不同原理。转运蛋白（底部）用一个反式二聚体表示，这提供向内和向外开口构象的一个简单的设计基础。当这与一个能量来源如 ATP 水解或者电化学梯度结合起来的时候，这个转运蛋白就能成为一个定向转运单元。通道蛋白和转运蛋白的动力学通常很不一样——分别受限于扩散速率和大的构象变化。

许多通道都是被门控的，这表明有信号来打开或者关闭它们。这个信号可以是如 cAMP、钙离子、光信号或者是膜电势的变化。这样一个被门控的通道同时也是一个信号转移体，或者换言之是一个受体蛋白。门控离子通道是亲离子受体。类似地，构象变化，如那些底物结合到转运蛋白上引起的构象变化，被用作跨膜信号转导的一种方式。一个

转运体系可以被简化为一个非转移的感知体系，当底物结合到感受器时引起一个构象变化。这个构象变化随后吸引胞内因子，并将之激活，引起下游反应。这类的例子包括 G 蛋白偶联受体（见第十一章），包括细菌视紫红质和一系列广泛分布的受体。当这类受体通过第二信使间接调节离子通道时，它们被称为促代谢受体而不是亲离子受体。

10.4.1 膜通道蛋白

10.4.1.1 膜孔蛋白

自从 1992 年得到第一个β桶型蛋白的结构以来(图 10.4),很多这类结构涌现了出来。其结构要么是用 X 射线晶体学，要么是用核磁共振方法解析得到，它们都是细菌外膜膜蛋白。有利的方面是，许多这类膜孔蛋白能在大肠杆菌中以包涵体的形式大量表达然后被重折叠，因此能够为核磁共振有效地合成放射性标记的蛋白质或者为晶体学有效地合成硒代甲硫氨酸衍生蛋白。

β 桶上条带的数目会影响核孔内径，而覆盖着或者填充着内部空间的结构单元可以形成门控机制里的一个盖子或者塞子，如在 TonB 依赖的 FecA 转运蛋白里观察到的一样（图 10.8）。

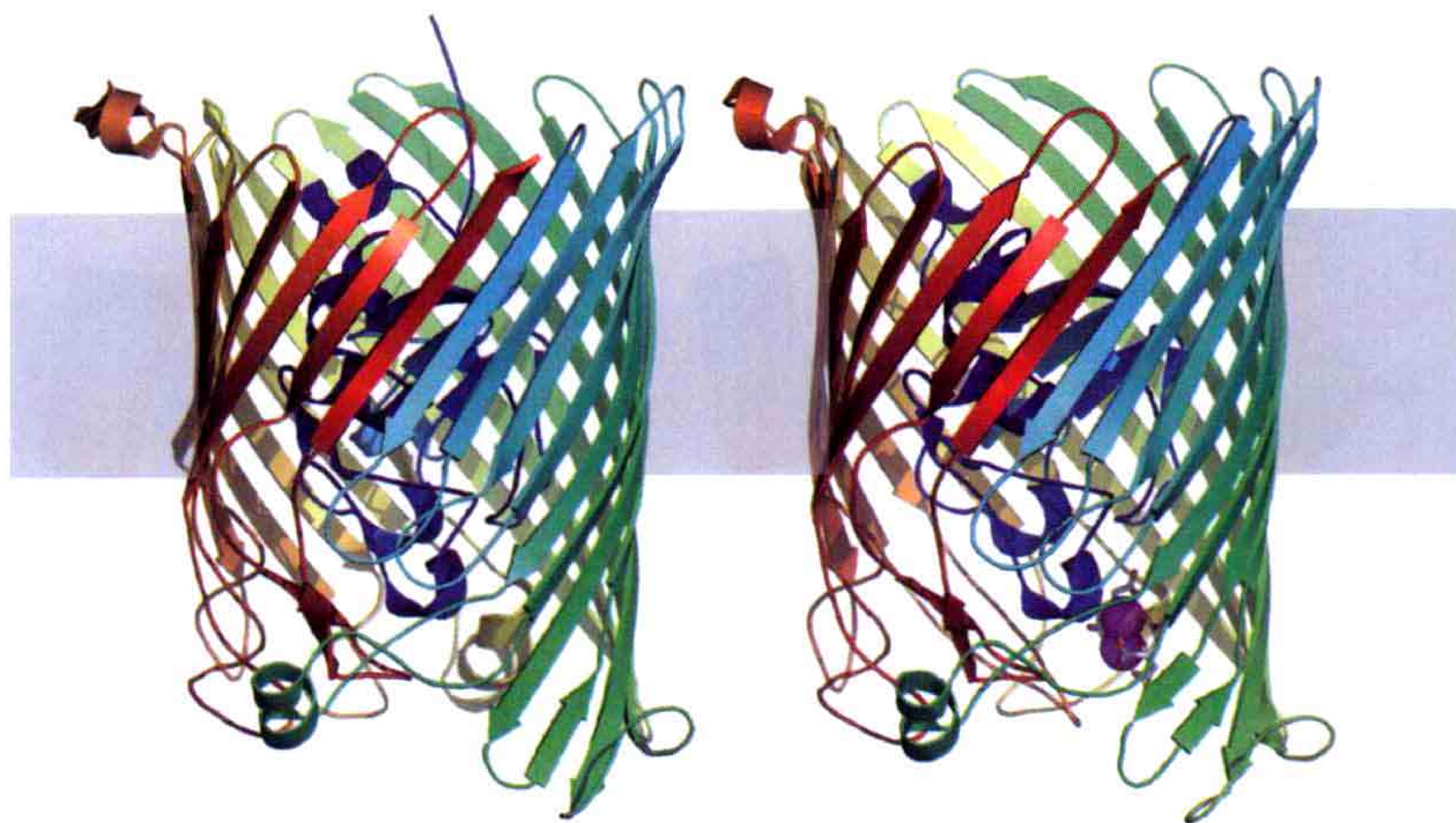


图 10.8 大肠杆菌外膜柠檬酸铁转运蛋白 FecA（PDB: 1KMO，1PO3）是基于一个 22 个片层的β桶型结构的（天蓝色到红色色谱）其 N 端结构域（蓝色）塞在中间作为两个柠檬酸螯合铁离子（右边底物结合复合体里面的白色小棍和紫色小球）的门控结构域。

β 桶型蛋白还是一些细菌和病毒致病性的基础，协助其入侵或者传播或利用自组装的膜孔穿透宿主细胞膜。一个研究较为透彻的例子是 α 溶血素膜孔，它是由可溶性的 α 溶血素在宿主细胞膜上寡聚化形成的。事实上，这个机制和先天免疫系统的一个关键组成部分，即攻膜复合物很像，它通过补体因子 C5b 和 C6、C7 和 C8 与几个 C9 因子形

成复合体，在病原和其他异源细胞（如移植器官的组织）的膜上形成一个大的跨膜膜孔。

许多膜孔蛋白，如 OmpC、OmpF 和 PhoE，都是革兰氏阳性菌的外膜微孔。它们与各自的底物结合时并没有显著的亲和性。它们允许小的极性底物（< 600Da）借助浓度梯度来通过。对于低浓度状态下的底物被动扩散就无法满足了。这样的底物通过底物特异的通道来运输，如对麦芽糖特异的 LamB、对磷酸特异的 PhoE 及运输柠檬酸铁的 FecA（图 10.8）。

10.4.1.2　水通道

水可以跨膜扩散，但是扩散是低效率的，不能满足很多细胞的生理需求。因此，很久以前就有人推测存在水通道，不过水通道直到 20 世纪 90 年代初才被 Peter Agre 发现。这类通道被称为水通道，它们是有许多相关亚型大量存在的蛋白质。许多细胞拥有这些对水分子有特异性的膜通道。在这个家族里面同样有对甘油有通透性的蛋白质。水通道蛋白解决了一个困难的选择性问题：在允许水通过的同时如何避免小的离子或者质子的通过?

水通道是四聚体，每个亚基都有孔道。蛋白质由 6 个跨膜螺旋和 2 个在膜中间“相遇”的短螺旋组成（图 10.9）。其 N 端半部分和 C 端半部分似乎起源于基因重复，并且有近似的二重轴相关对称性。通道具有沙漏状结构，两个门庭由一个 20Å 的通道相连，通道最窄部分的宽度不超过 2.8Å。水分子需要排成单列穿过这个通道。对称的两个 loop 都包含高度保守的标志性模体，即 NPA，两个 NPA 序列并排位于通道上。

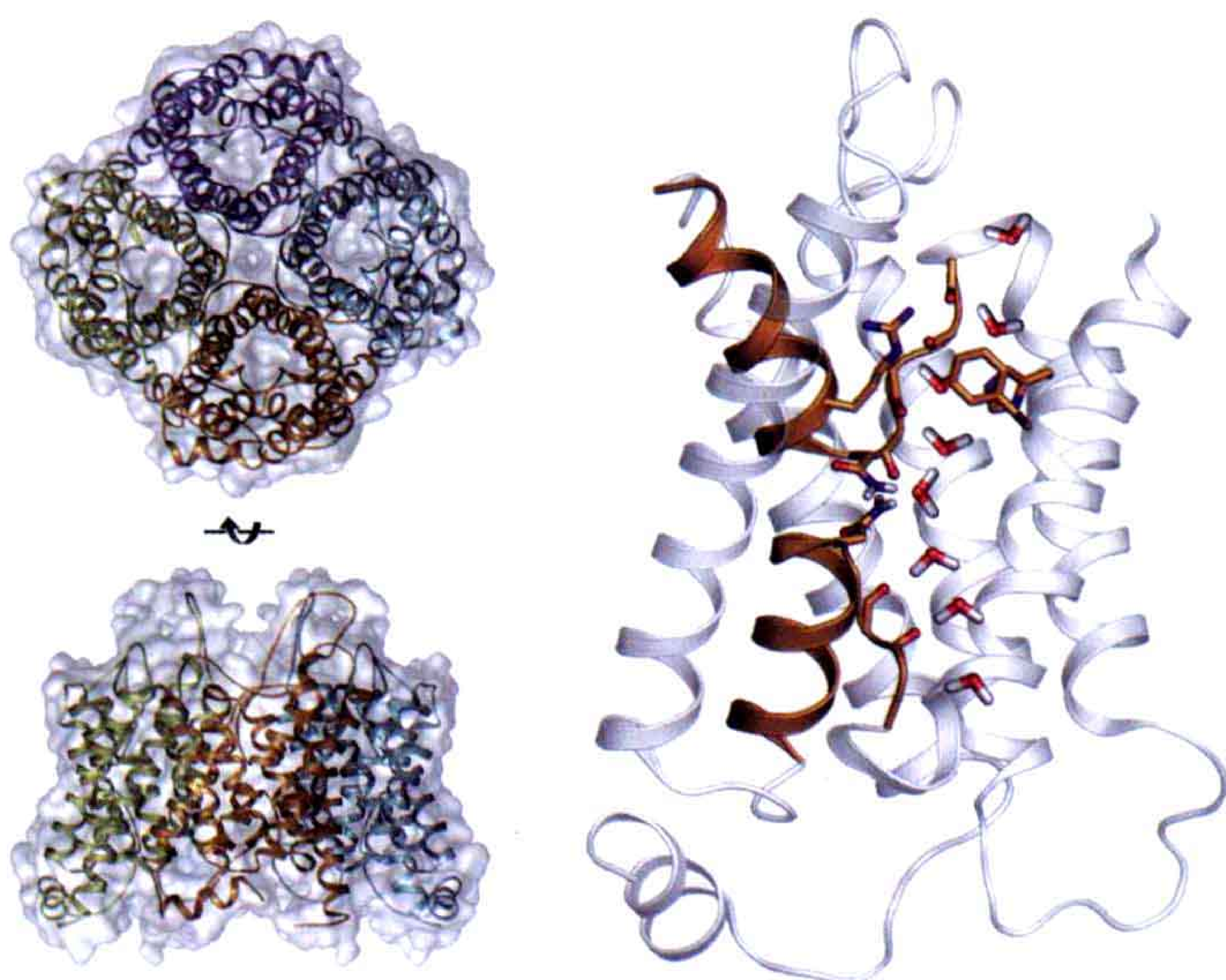


图 10.9　水通道四聚体（左，上图和下图）在每个单体里面都有一个孔道（PDB：1J4N）。右：从平行于膜表面的方向看到的水通道单体。这个分子围绕着图形中间的轴有着近似的二重对称性，对称轴垂直于图中心与纸平面平行。图中还显示了保守的 NPA 模体上的天冬酰胺及限制水合氢离子通过的精氨酸。不同朝向的水分子通路是基于分子动态模拟研究画出的。

孔道具有部分疏水的表面，但是也有一些极性基团包括一些羧基氧，这允许水分子通过孔道。在约束区，水分子可以和 NPA 标志模体上保守的天冬酰胺形成氢键。而由于孔道内存在一个精氨酸带正电荷，两个短的螺旋的 N 端有部分正电荷，以及没有一些合适的配体来补偿水合外层缺失，带电荷离子，主要为 H_3O^+，会被阻止不能穿过孔道。分子动态模拟进一步指出一个基于孔道中心水分子旋转的重要选择机制，这个又一次说明对水分子具有选择性而不是水合氢离子。

10.4.1.3 钾通道

钾通道既存在于真核细胞又存在于原核细胞。例如，在真核细胞里，它们在神经元中为电压门控通道。这些通道根据电压来调节开关，对于神经传导中的动作电位形成具有极重要的作用。有些通道也可以被其他信号，如钙离子的结合或 cAMP 等来调控，机械感应同样可以作为控制开关的一个门控机制。总的来说，通道对于特异离子如钠、钾离子呈现出选择性，反之亦然。

钾通道是一类四聚体分子家族的成员，这个家族还包括其他阳离子通道如钠离子通道和钙离子通道。它们都有不同数量的跨膜螺旋和由两个跨膜螺旋间的片段组成的孔道。一个细菌的钾通道，即 KcsA，只有两个跨膜螺旋，在四聚体里构成一个有 8 个螺旋的螺旋束。这个通道的结构已经被解析出来了。4 个亚基绕垂直于膜的轴有四重对称性排列（图 10.10）。与每个单体含有一个通道的水通道不一样的是，孔道形成于 4 个亚基的中间。四聚体具有一个选择性滤器、一个中央门庭和一个门控区域。

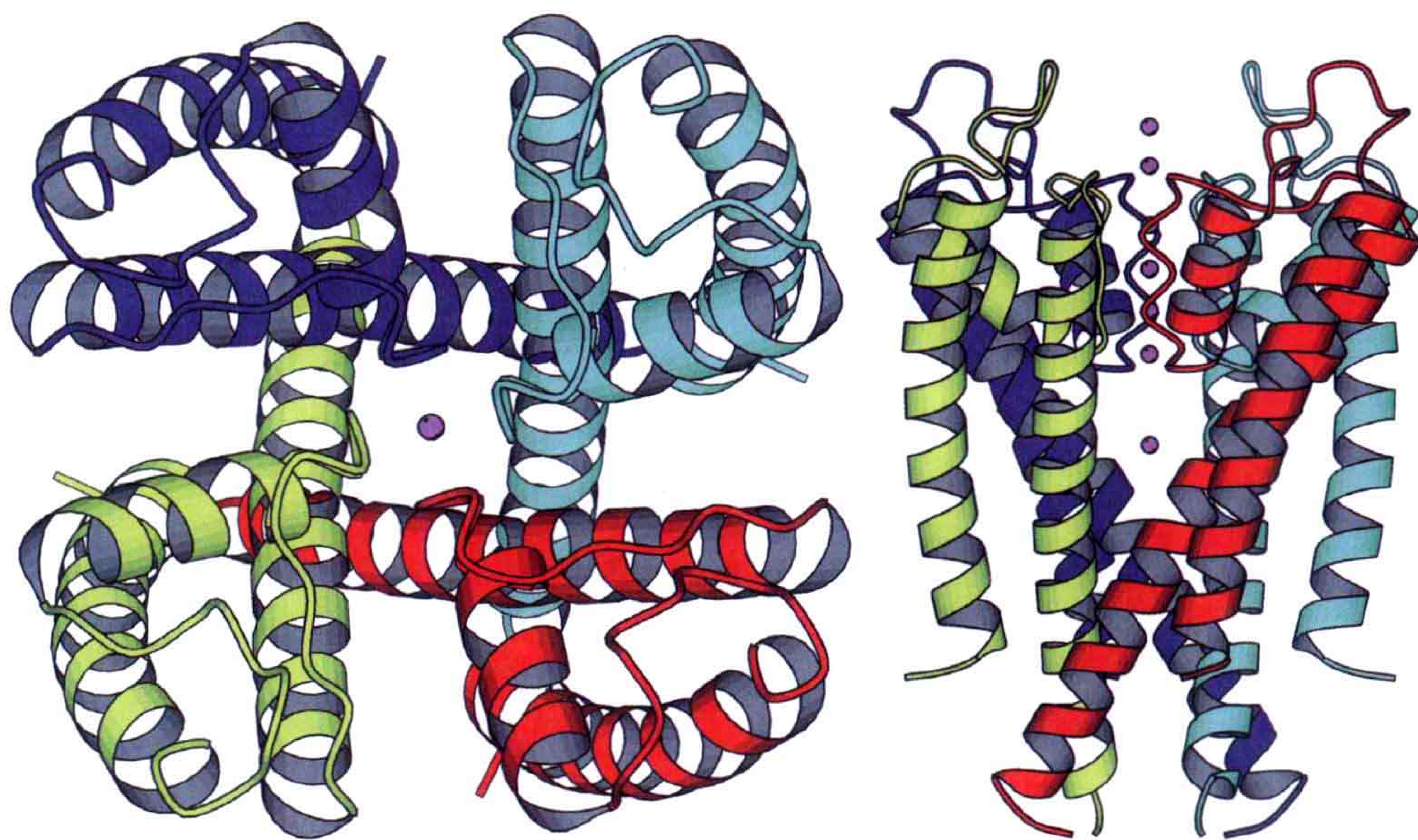


图 10.10 KscA 钾通道。从上看四聚体（左）及从侧面看四聚体（右）。四聚体具有一个选择性滤器和一个膜内的中心通道，通道由构成滤器的螺旋的双极性稳定住，这一点大幅减小了实际跨膜距离。下端通路的构象决定门控是打开还是关闭（PDB：1K4C）。

在外侧表面附近，4 个亚基的 loop 环形成一个孔道，这个孔道有选择性过滤器的作用。这个选择性过滤器包含一些指向通道中间的羧基氧（图 10.11）。这个属性解释了孔道是如何对钾离子有特异性，而阻止小一点的钠离子高效通过的。其原因在于氧原子间的精确距离。在这个滤器里，这个距离适合钾离子的结合。水合钾离子能以较低的能量损耗随之离开它们的水合层，然后利用与滤器上 4 圈羧基氧连续的相互作用来通过滤器。另外羧基氧之间的距离对于较小的钠离子来说太长了，这样钠离子就不能离开它的水合层，不能通过过滤器。利用主链氧来形成过滤器，较之利用侧链原子能够得到更加稳定的构象。这对构建一个能够区分钠离子和钾离子的通道可能是很关键的。通道的开关通过跨膜螺旋控制通往中央门庭的自由路径来保证。在钾离子浓度较低的情况下，这个过滤器呈现出一个更为紧闭的结构。

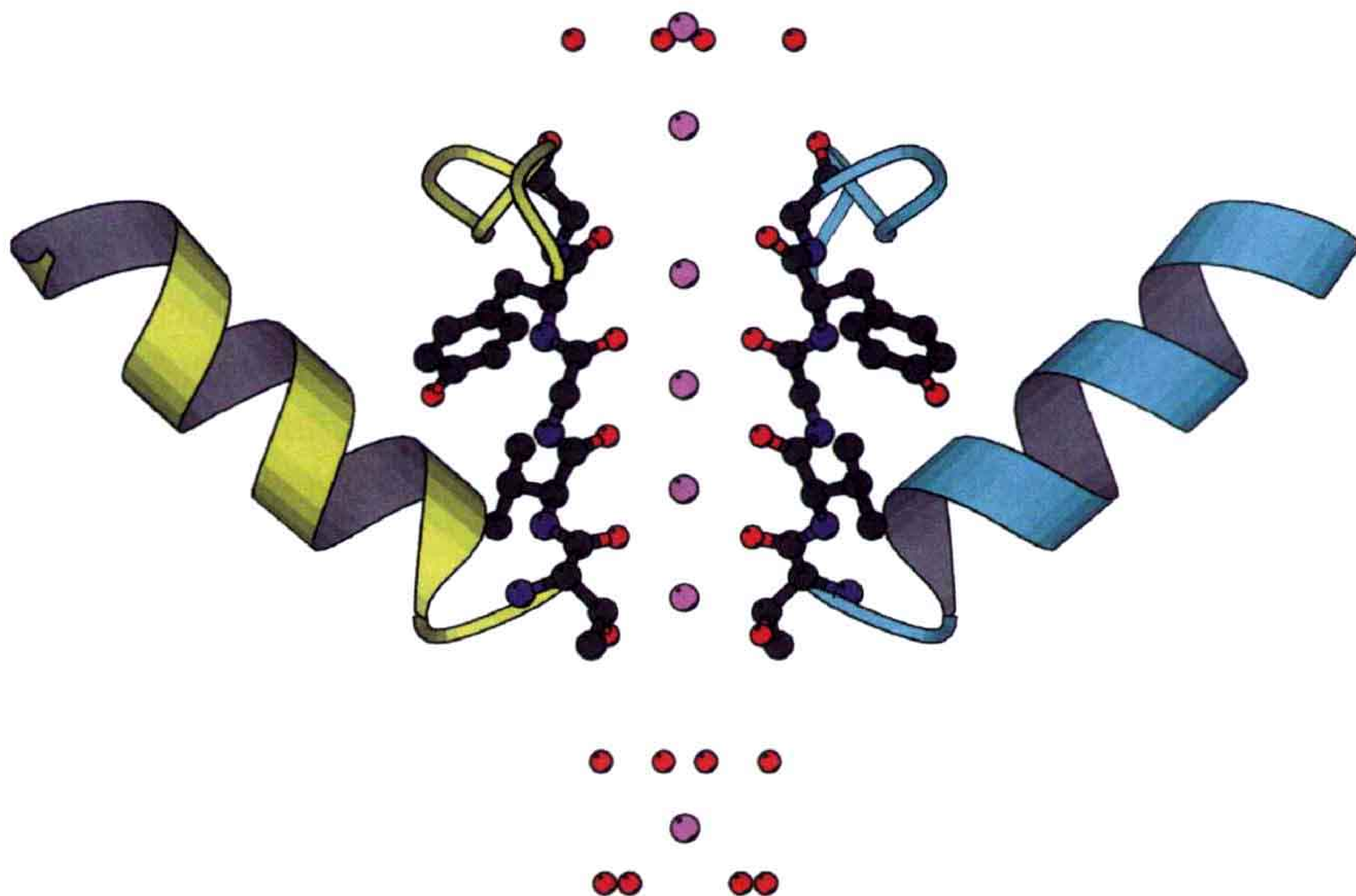


图 10.11　高浓度钾离子条件下 KcsA 的选择性过滤器。图中只画出了两个亚基。一些钾离子（淡紫色）正填充着滤器，但是在过滤器里面，只有每隔一个位置才可以同时被离子占据。羧基氧面向通道，限制离子通过，只有大小吻合滤器里面四聚体羧基间距离的离子才可以通过。在滤器下方的门庭里面有一个离子，被 8 个水分子结合着，并且被 4 个螺旋（图中画出了其中 2 个）的双极性负电荷稳定着。

10.4.1.4　钾通道的门控

有不少结构学研究试图阐明通道是如何被门控的。理解电压依赖门控机制的基础具有很重要的意义。门控机制依赖于电压感受器的运动，电压感受器含有一些精氨酸，这些精氨酸为了响应膜电势，通过膜带负电荷的一边袭击正电荷的方式在膜中发生显著运动，从而实现电压感受器的运动。基于从 *Aeropyrum pernix* 的电压门控通道结构（KvAP）和兔子的电压门控通道结构（Kv1.2）提出了 paddle 机制，精氨酸残基跨过膜是很难的。另外，这个机理提供了只有当给定电压达到给定阈值才有反应的开关机制的基础（图 10.12）。

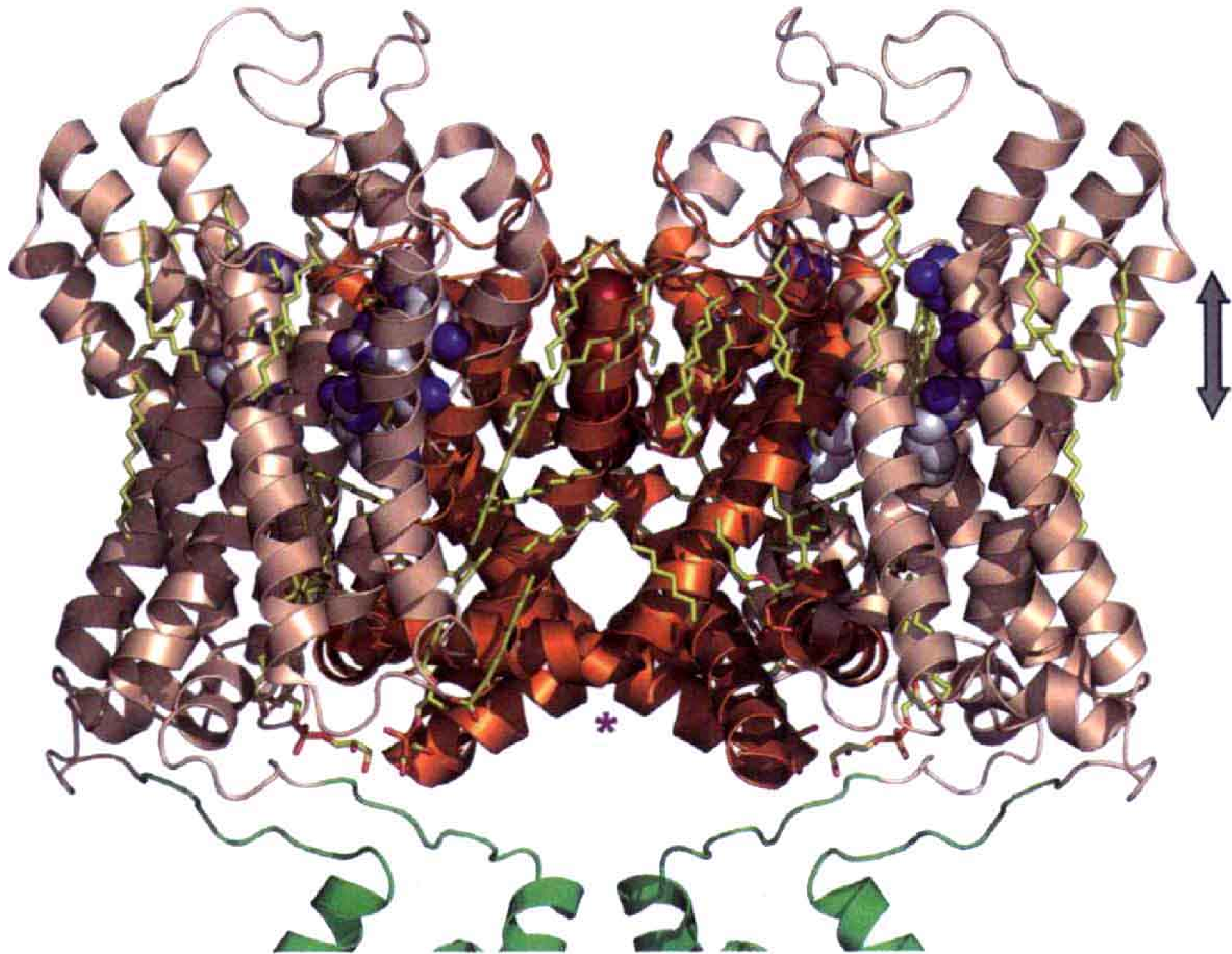


图 10.12　来自老鼠的电压门控通道 Kv1.2（PDB：2R9R）。这个四聚体的通道的核心区域用橙色表示，钾离子的位置用深红色小球表示。每个亚基有一个电压感知结构域，图中用浅棕色表示，参与电压感知的精氨酸残基用白色和蓝色（碳和氮）小球表示。这个感知结构域和绿色的胞内结构域相连接。只是膜双层的脂分子用黄色小棍显示。膜电位的变化会影响感知结构域在膜内的位置（箭头），然后控制通道的开关（星号）。

另一个重要的机制是配体介导的门控机制，如 MthK，它是一个钙门控的钾通道。该门控就是在胞质区的一个结构。它由 8 个相同的片段组成，其中 4 个连到组成孔道的 4 个结构域的 C 端（图 10.13）。这 8 个片段组成一个大环。结构域和结构域之间的相互作用是有点柔性的，依赖于与钙离子的结合。钙离子的存在影响门控环的大小，进而影响孔道在胞内开口的大小。

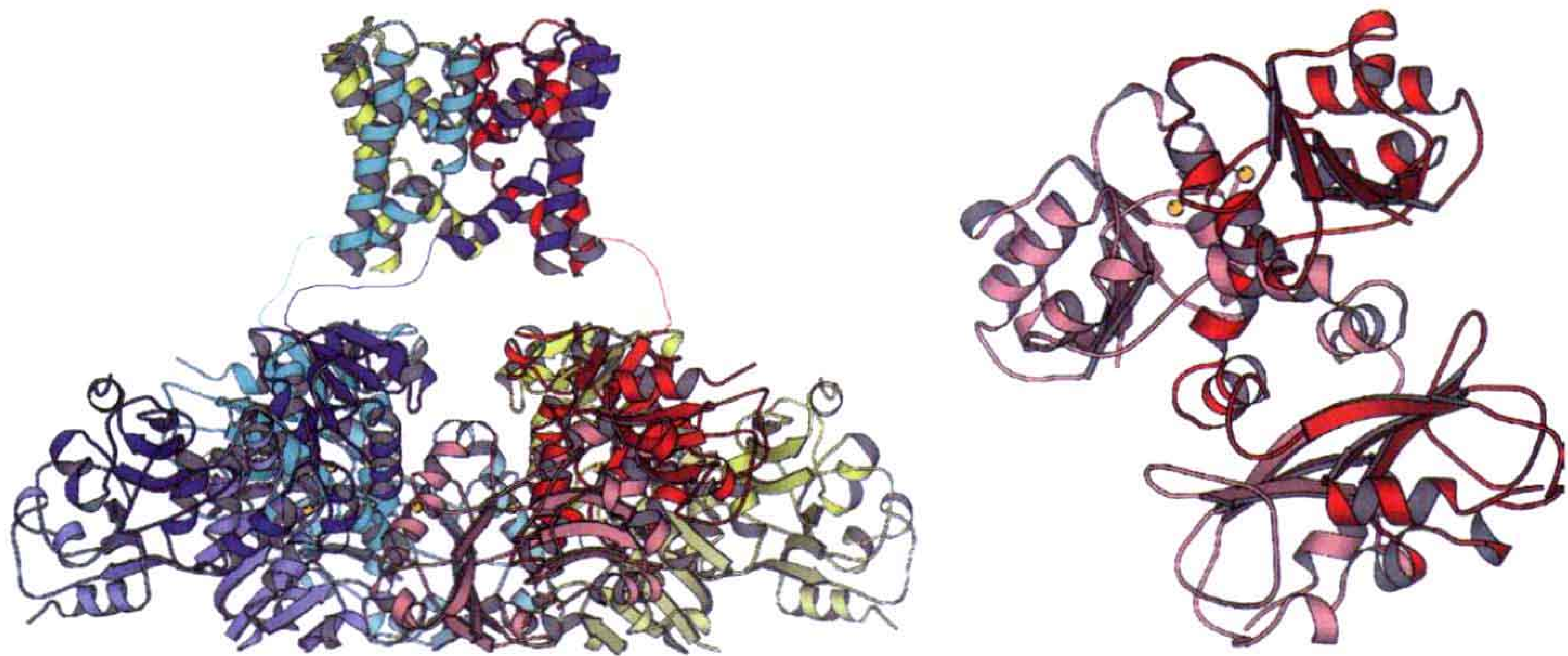


图 10.13　钙门控的钾通道。门控环的构象变化会通过与螺旋的连接导致孔道入口的开或者关。左：完整的八聚体分子（PDB：1LNQ）。顶部的滤器由 4 个亚基组成，和图 10.10 中的滤器相似。门控环由所有 8 个亚基的 N 端部分组成。右：以垂直于左图视角的方向看到的门控环里面的两个亚基。二重轴在画面的平面上，二聚体间的相互作用是由钙离子（黄色）来稳定的。

10.4.1.5　镁离子通道 CorA

镁离子对于细胞内许多过程的正常进行十分重要。对镁离子摄入机理的研究还不透彻，不过两类细菌的镁离子通道已经被解析出来了。其中一个，即 CorA，是真细菌和古细菌中的主要镁离子通道，它还和线粒体中的通道相关。与钾通道四聚体不同的是，这个蛋白质是一个五聚体。其单体的 N 端胞内部分主要由反向 β 片层构成，片层的两面都有螺旋。其中一面的螺旋比片层要长很多，其螺旋的 C 端跨过脂双层，加上另一个跨膜螺旋就构成完整的蛋白质（图 10.14）。这些螺旋的第一个跨膜螺旋组装成膜孔。蛋白质的胞内区构成一个漏斗，漏斗的窄口连接着膜孔开口。膜孔通道比跨膜片段要长，在跨膜部分附近具有三圈疏水残基（Met291、Leu294 和 Met302），构成膜孔通道的限制区。疏水限制区构成离子通透的一个很强的能量障碍，这可能也是 CorA 转运离子速度很慢的部分原因。

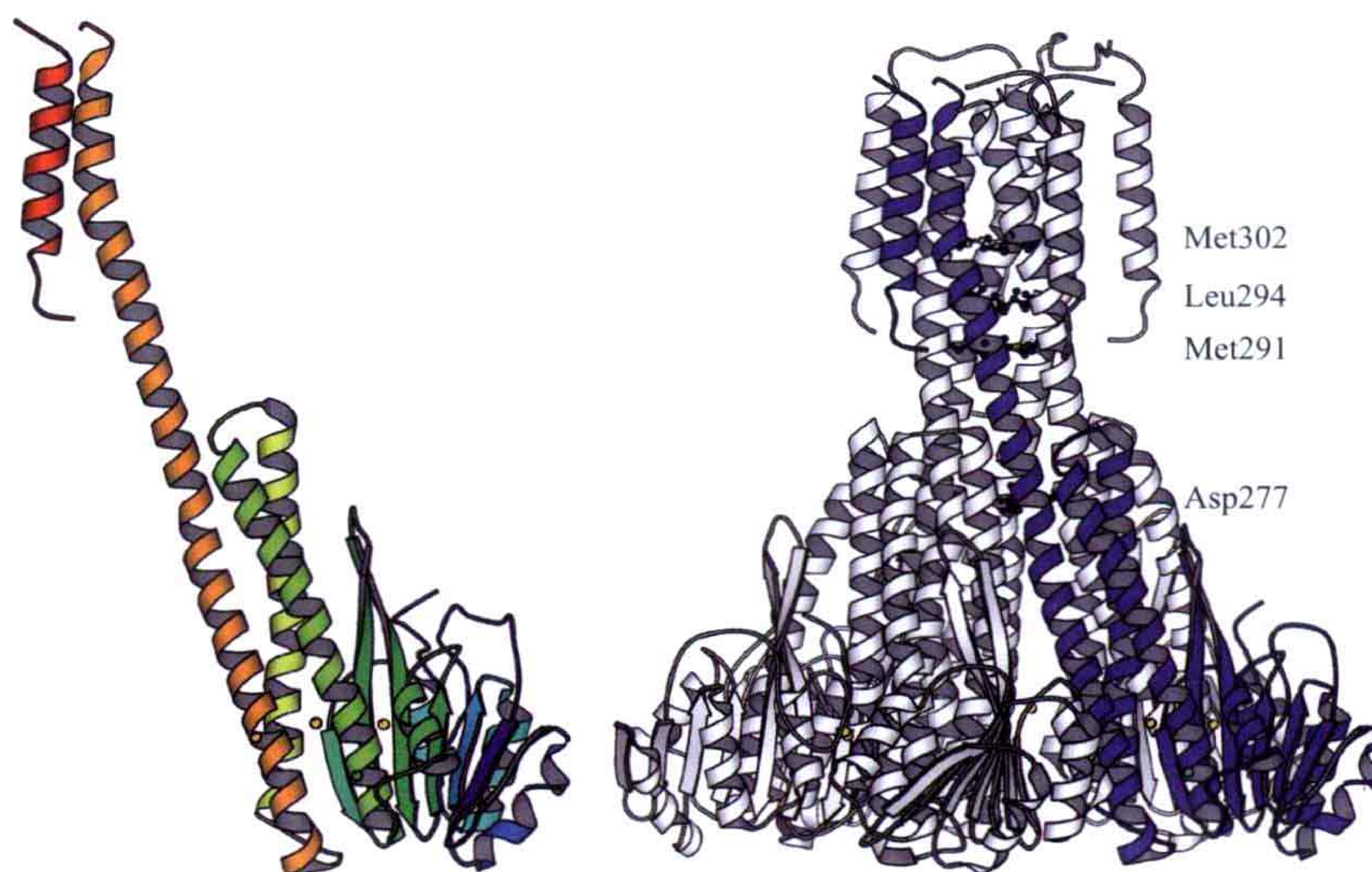


图 10.14　CorA 蛋白。左：一个颜色从蓝色（N 端）到红色（C 端）标记的单体。结合在亚基上的两个镁离子（金色）和一个氯离子（绿色）的位置用球体标记出来。右：五聚体复合物，其中一个亚基用蓝色标记。嵌入膜的膜孔在上面，在膜孔一边有蛋白质胞内区构成的一个漏斗状的结构。构成膜孔表面的三个疏水残基的位置及连接漏斗和膜孔的天冬氨酸残基的位置都在图中被标记出来了（PDB：2IUB）。

镁离子需要严格的八面体配位方式。这样需要一个正方平面的 4 个配体及上下各一个配体来完成配位。选择筛的五重对称性导致的对称性不符也是蛋白质对镁离子转运速度很慢的原因之一。CorA 转运镁离子的结构细节尚不明确，不过可能涉及一些构象变化。

在胞内区亚基界面上发现了两个镁离子结合位点。这些位点可能通过感知镁离子浓度来调控通道的属性。

10.4.1.6　氯离子通道或者 Cl^-/H^+ 反转运蛋白

从细菌到哺乳动物中都存在一类大的氯离子通道蛋白（ClC）家族。人体中这类

蛋白质的缺陷导致一系列疾病，如囊性纤维性病变。一些 ClC 通道蛋白具有调节氯离子流动的门控结构域，如导致囊性纤维性病变的蛋白质，这个蛋白质是以 ATP 结合匣（ABC）模式被门控的。

有些 ClC 蛋白被证明是 H^+/Cl^- 反转运体，也就是说更像次级转运蛋白而不是通道蛋白，虽然通道蛋白和转运蛋白本来就有很大的相似性。这样，在人体中，定位于核内质溶酶体膜上的 ClC 蛋白是反转运体，它们和转运质子的 V 型 ATP 酶合作，而定位在质膜上的 ClC 蛋白则是通道。这些发现刺激了对转运过程和门控通道过程不确定性的研究，也刺激了对于利用反转运体替换通道（如在溶酶体腔酸化过程中发生的）导致的生理后果的研究。

一个细菌 ClC 蛋白的结构已经被解析出来，该蛋白质是一个 Cl^-/H^+ 反转运体（图 10.15）。这个反转运体是由两个三角形亚基组成的二聚体，每个亚基包含一个完整的通道，或者说转运途径。每个亚基由 18 个 α 螺旋组成一个复杂的拓扑结构。大部分螺旋不与膜面垂直，有些螺旋并不完全穿膜。这个亚基是由 N 端和 C 端有相似折叠方式的两半组成，因此每个亚基绕着膜面的一个赝对称轴展示出二重对称性，表现出倒转对称的拓扑模型。通过仔细的基于结构的序列比对发现两端序列有很弱的相似性，这提示可能是由基因重复造成的。这也表明 ClC 蛋白可能起源于一个转运蛋白系统（见 10.6 节）。

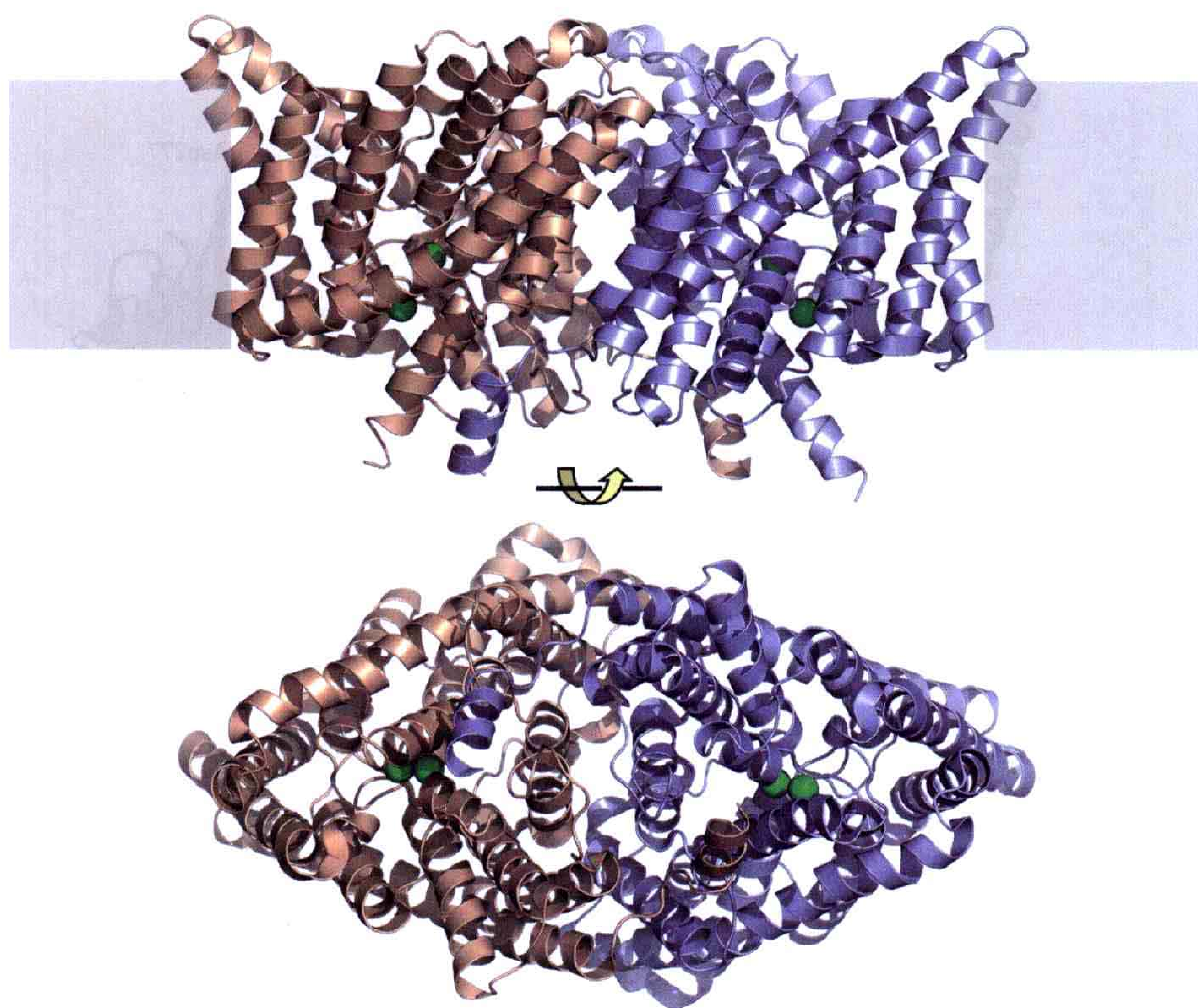


图 10.15 来自大肠杆菌的 ClC 家族氯离子通道（Cl^-/H^+ 反转运体）二聚体（PDB：1OTS）的侧面图（上）和从胞内视角所见示意图（下）。两个亚基分别用浅橙色和浅蓝色卡通标识，氯离子用绿色小球标记。每个亚基构成一个有功能的通道 / 转运体，在中心位点和胞内入口上各结合了一个氯离子。

选择性滤器位于分子中央（图 10.16）。由 4 个 loop 组成的中心位点上有一个保守的酪氨酸和丝氨酸残基与氯离子相互作用，这些残基都靠近或者位于螺旋的 N 端。α 螺旋有双极性，螺旋 N 端的部分正电性对于在膜中间稳定住负电荷可能是很重要的。这个机理在许多结合阴离子的蛋白质中都存在。相似地，阳离子通道则利用螺旋的 C 端在膜中间稳定正电荷（见 2.1.4.3 节和 10.4.1.3 节）。

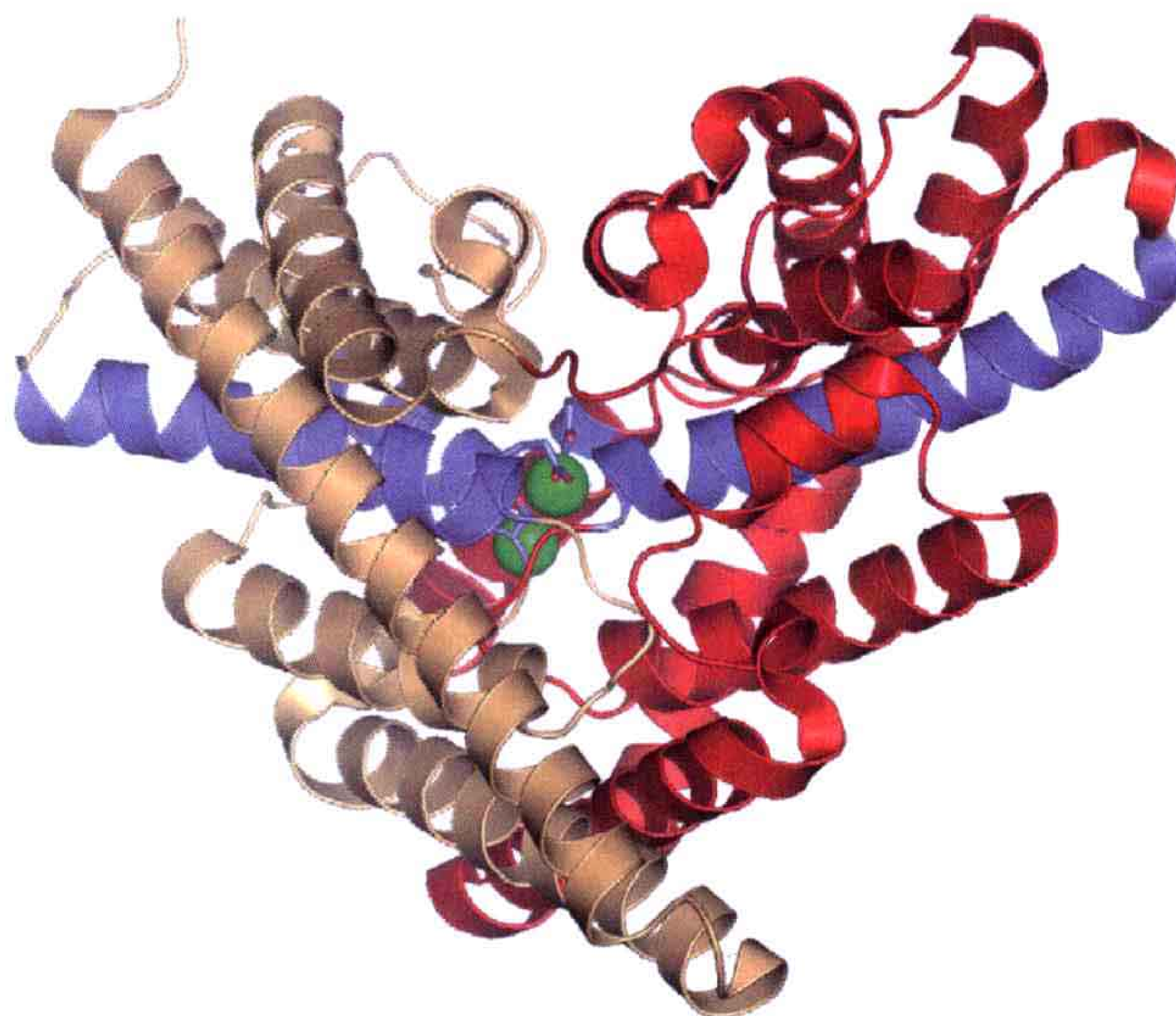


图 10.16　从胞外视角看到的一个单独的 ClC 亚基。其展示了分子内的赝对称性（浅橙色和红色）。中心位点被 α 螺旋的 N 端稳定住（蓝色）。定位于中心位点上的一个谷氨酸是一个门控残基，它也有可能参与 ClC 反转运体的质子偶联转运过程。

10.5　初级转运蛋白

初级转运蛋白利用外源能量如光或者 ATP 来主动跨膜转运。这使细菌视紫红质可以转运质子；光系统转运质子和电子；V- 型和 P- 型 ATP 酶泵转运如质子、钠离子、钾离子和钙离子之类的阳离子；还有 ABC 转运蛋白可以转运更大的化合物。在 Sec.5.3.2 中介绍了负责合成 ATP 的膜上动力蛋白——细菌、叶绿体和线粒体中的 ATP 合成酶。

10.5.1　光驱动的质子转运——细菌视紫红质

细菌视紫红质是一个复杂而又简单的质子转运系统，在这里利用一个视网膜发色团来吸收光子的能量，而这个发色团是通过和 Lys216 有席夫碱（Schiff’s base）式连接而共价结合到蛋白质上的（图 10.17）。吸收光子刺激了视网醛从全反式基态到 13- 顺式高能态的异构化。通过和 Lys216 的共价连接，这个光异构化作用会驱动蛋白质进行

构象变化，进而有效地把席夫碱从和 Asp85 的离子相互作用中分开。膜内有单独的电荷是一个很不稳定的状态，而质子从席夫碱转移到 Asp85 的侧链上（对应着 pKa 值的变化）导致了电荷中和。在从高能态缓解下来的过程中，Arg82 会靠近 Asp85，诱导去质子化（重设 pK_a）。质子再通过 Glu194-Glu204 对释放到胞外环境中。而席夫碱会通过一个被包被的 Glu96 残基从胞质区重新质子化。这样就构成了一个每一个光异构循环都从胞内向胞外转运一个质子的质子转运链。这样的机理 —— 一个含有用一个精氨酸残基调节 pK_a 值的可滴定的天冬氨酸和谷氨酸残基的质子转运通道——存在于许多质子转运系统中。

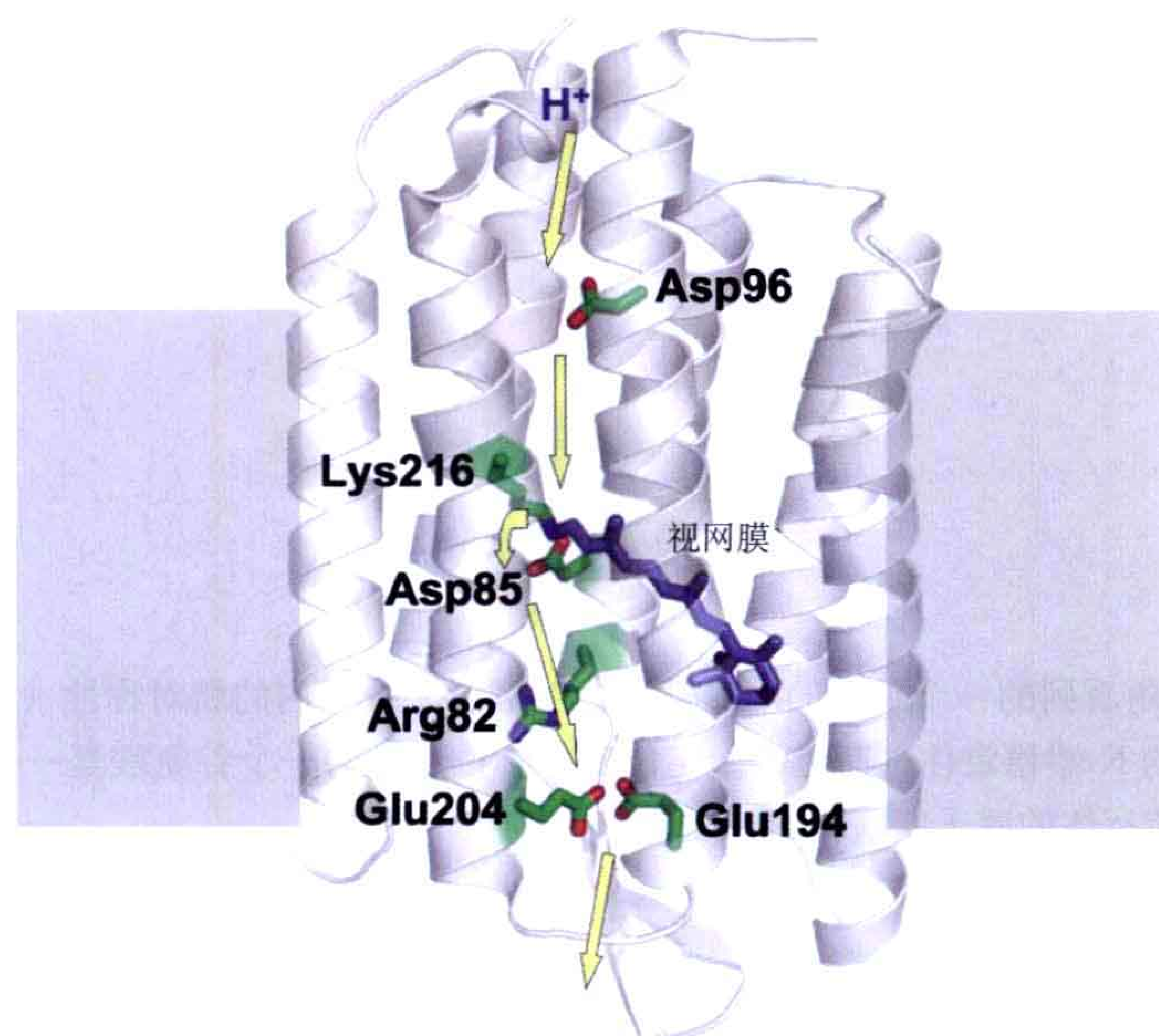


图 10.17 细菌视紫红质的结构，侧链和箭头指示了质子转移的通道。质子转移和以席夫碱连接于 Lys216 侧链氨的视网醛的光驱动的顺势 - 反式异构化相偶联。

10.5.2 光驱动的氧化还原过程——光系统

生物圈的能量归根到底来源于太阳，利用日光的能量把水分解成氧气的能力是演化中的一个重要转折点，从此大气层逐渐转变成一个氧化性的环境。光合作用利用了大概 100TW 的能量，这大概占了所有照到地球上的太阳能的 0.1%。通过光合作用，植物和微生物从空气中吸收二氧化碳生成葡萄糖。

光合作用基于两个光系统，即光系统 I （图 10.18）和光系统 II，它们的最大吸收波长分别是 700nm 和 680nm。光吸收是基于由“特殊对”稳定住的电子激发和电荷分离。“特殊对”即两个叶绿素环，每个都通过电子传递链连接到电子受体上。在光系统 II 中，这是一个分解水的锰催化中心，能制造氧气。

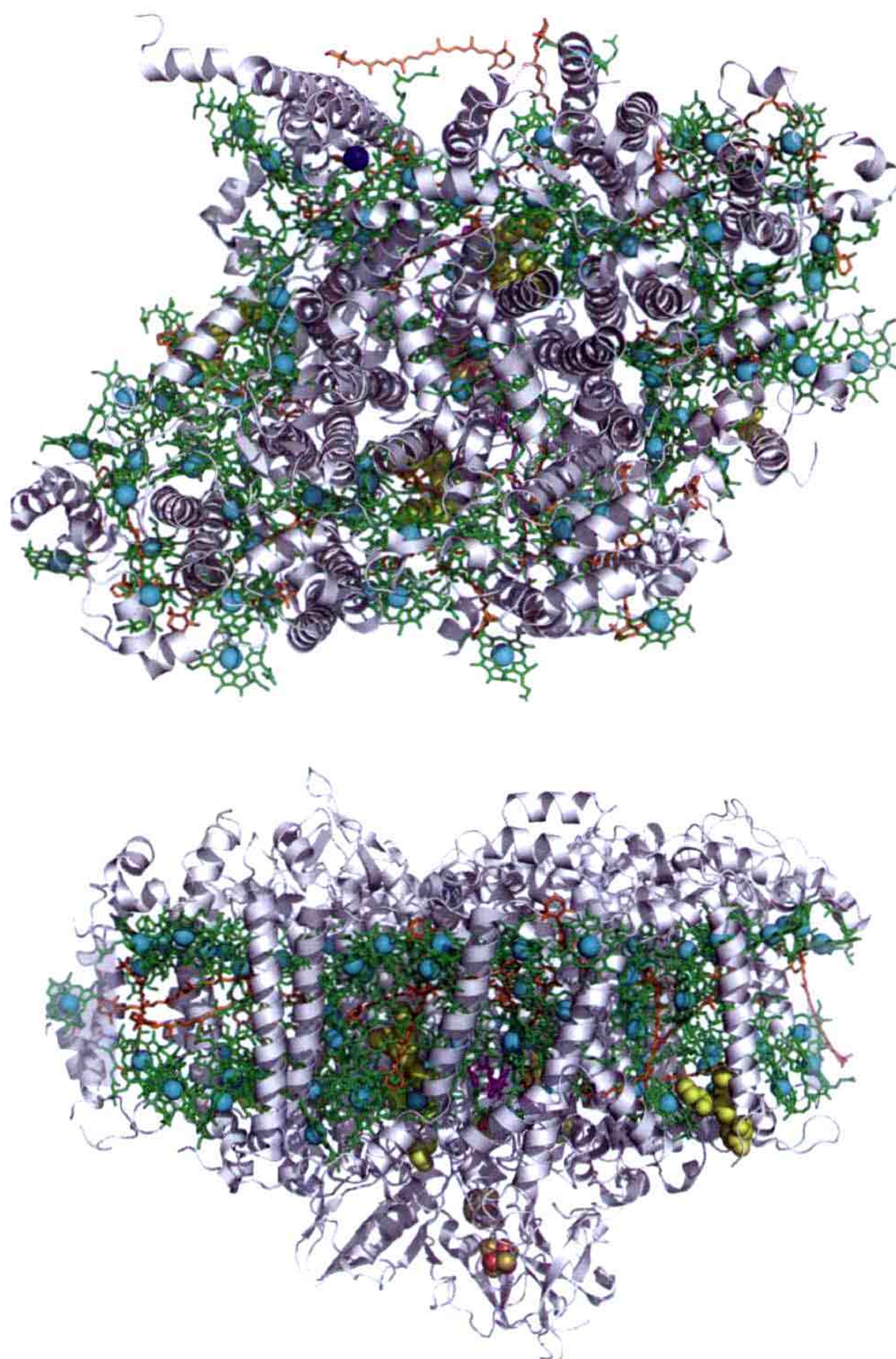


图 10.18　来自 *Thermosynchococcus elongates* 的光系统 I 的结构。上：从类囊体内腔朝膜看到的结构；下：从类囊体膜侧面看到的结构。这里一共有 12 个蛋白质亚基，其中两个（A 和 B）有 11 个跨膜螺旋，构成蛋白质的核心。单体结合 96 个叶绿素、3 个 Fe_4S_4 铁硫簇、2 个质体醌，22 个胡萝卜素和 4 个脂分子。蛋白质亚基用白色卡通指示出来，配体用棍状或者球状展示，叶绿素被标记为绿色（镁离子为天蓝色），β - 胡萝卜素为橙色，脂为黄色，质体醌为紫红色，Fe_4S_4 铁硫簇为橙色和黄色小球，钙离子为蓝色。注意围绕着特殊对（上图的中心）的赝对称性及被深深包被在内的脂分子作为辅基存在。这个结构最完整地展示了一个大型的指引激发态电子进入电子转移链的天线系统，这个电子转运链导致了电荷分离和氧还电势的建立（PDB：1JBO）。

这个反应事实上要远较上述复杂，涉及一系列电子传递，从而把光激发电子的高度活化的还原电势转化为由水到氧气的还原反应，同时造成质子梯度。

对于如何调节特殊对吸收光子的最优波长，目前已有进展（图 10.19）。通过突变 *R.viridis* 的光合反应中心 L 链 168 位组氨酸为苯丙氨酸，可以观察到显著的蓝移和初始电子传递速度的增加。His168 扮演着重要的角色，它是特殊对的氢键供体，把特殊对稳定在极化状态。将 His168 突变成为苯丙氨酸去除了对特殊对极化状态的稳定作用，从而需要更多能量来激发反应。

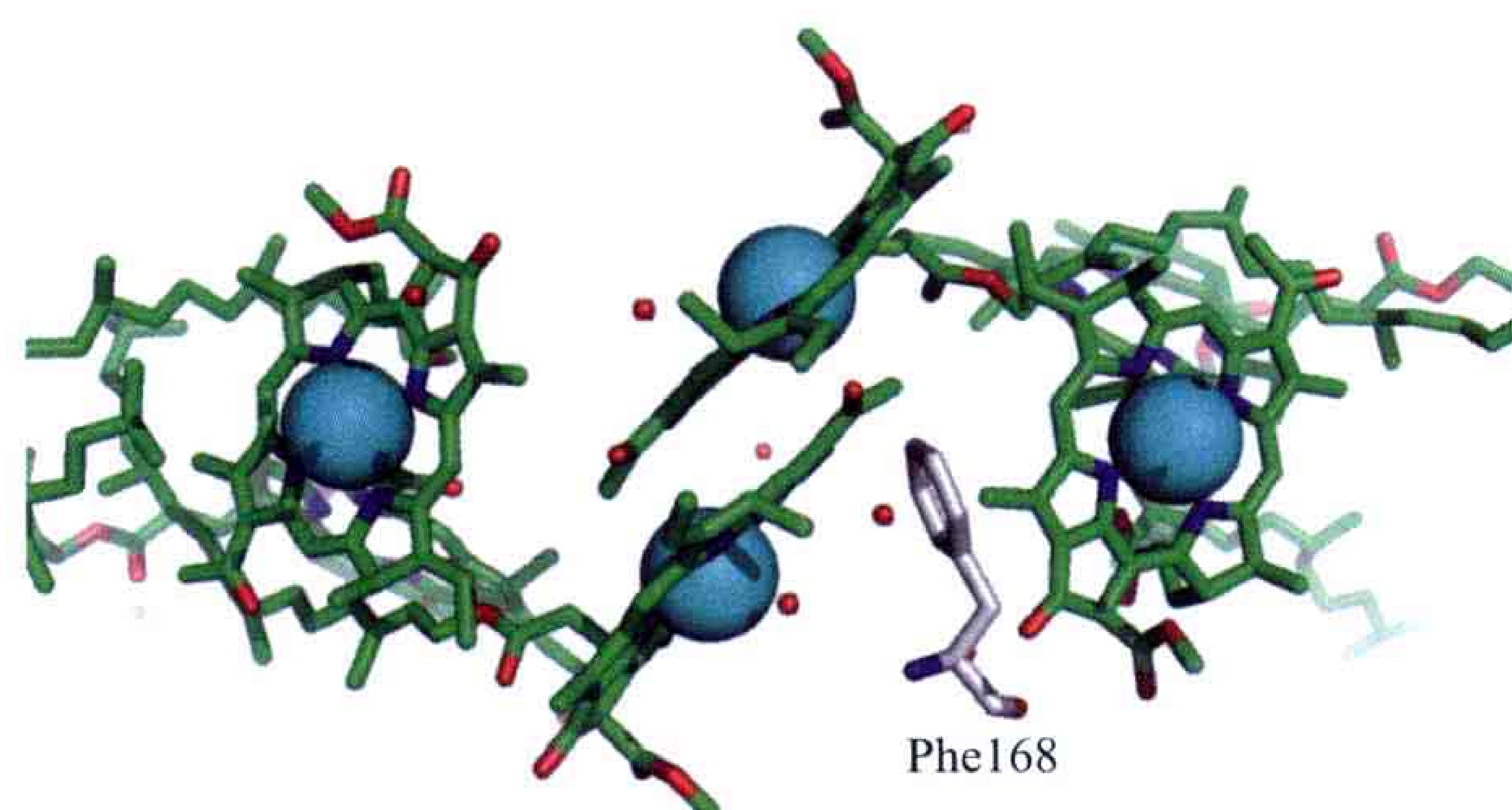


图 10.19 来自 *R.viridis* 的光合反应中心 L 链 168 位组氨酸突变成苯丙氨酸突变体上的特殊对显示出显著的蓝移及初始电子传替速率的增加。His168（位置用白色棍状的 Phe168 指示）和特殊对（绿色小棍，镁离子用天蓝色小球表示）相互作用。苯丙氨酸侧链对于极性的特殊对没有稳定作用（PDB：1XDR）。

10.5.3 ATP 驱动的泵，P 型 ATP 酶

P 型 ATP 酶家族的阳离子泵造成了生物膜的跨膜电化学梯度，并维持着细胞内的阳离子浓度和渗透压。它们把 ATP 的化学能以阳离子梯度，最终是膜电势的方式转化为电化学势能。家族成员包括钠钾泵、氢钾泵和钙泵等，它们都对生理过程起很关键的作用。P 型 ATP 酶消耗体内 1/3 的 ATP 这一现象也不足为奇。然而，真菌和植物的质膜也是由一种 P 型 ATP 酶供能的，这种质子泵造成强电势和高 pH 梯度，从而从酸化的介质中吸收营养物质。Skou 在 1957 年鉴定钠钾泵成为 P 型 ATP 酶家族的第一个成员。一种相似的钙离子激活的 ATP 酶在肌肉细胞储存钙的膜（肌浆网）中被发现。在 20 世纪 70 年代，一个 ATP 偶联的质子转运蛋白在脉胞菌质膜中被发现。后来发现这个质子泵与钠钾泵及钙泵是相关的，从而确认了普遍存在一类有磷酸化酶中间过程的蛋白质。

P 型 ATP 酶利用一个共价的天冬氨酸 - 磷酸中间产物的形成和断裂，与主动向胞外转运离子和向胞内转运反向离子的过程相偶联。P 型 ATP 酶的第一个结构在 2000 年发表，那是肌浆网钙泵（SERCA）不结合核酸并被钙离子稳定住的状态下的结构。这个结构展示了胞质区由 3 个结构域组成的头部和一个由 10 个跨膜螺旋，即 M1 到 M10 组成的跨膜结构域。这个结构与序列分析及许多突变实验推测的拓扑结构和钙离子结合残基位于跨膜螺旋 M4、M5、M6 及 M8 上等结果几乎完全吻合。此外，这个结果也和早期电镜研究结果吻合，虽然它的胞质区结构域是处于一个不常见的构象。胞质区结构域共有 3 个，分别称为 P（磷酸化）、N（核酸结合）和 A（传动）结构域，它们在结构上和 M1 ~ M5 相连（图 10.20）。它们为 ATP 水解和阳离子转运的有效偶联提供一个明确的基础。正如推测的，膜中间两个位点的对钙离子的高亲和力结合是以对钙离子正确的八面体配位为基础的。不过最值得注意的是两个钙离子结合位点的协同基础是连接两个位点的天冬氨酸的侧链。在钾离子通道中（见 10.4.1.3 节），主链羧基参与阳离子结合。这也是干扰早期用突变方法鉴定结合位点的原因。

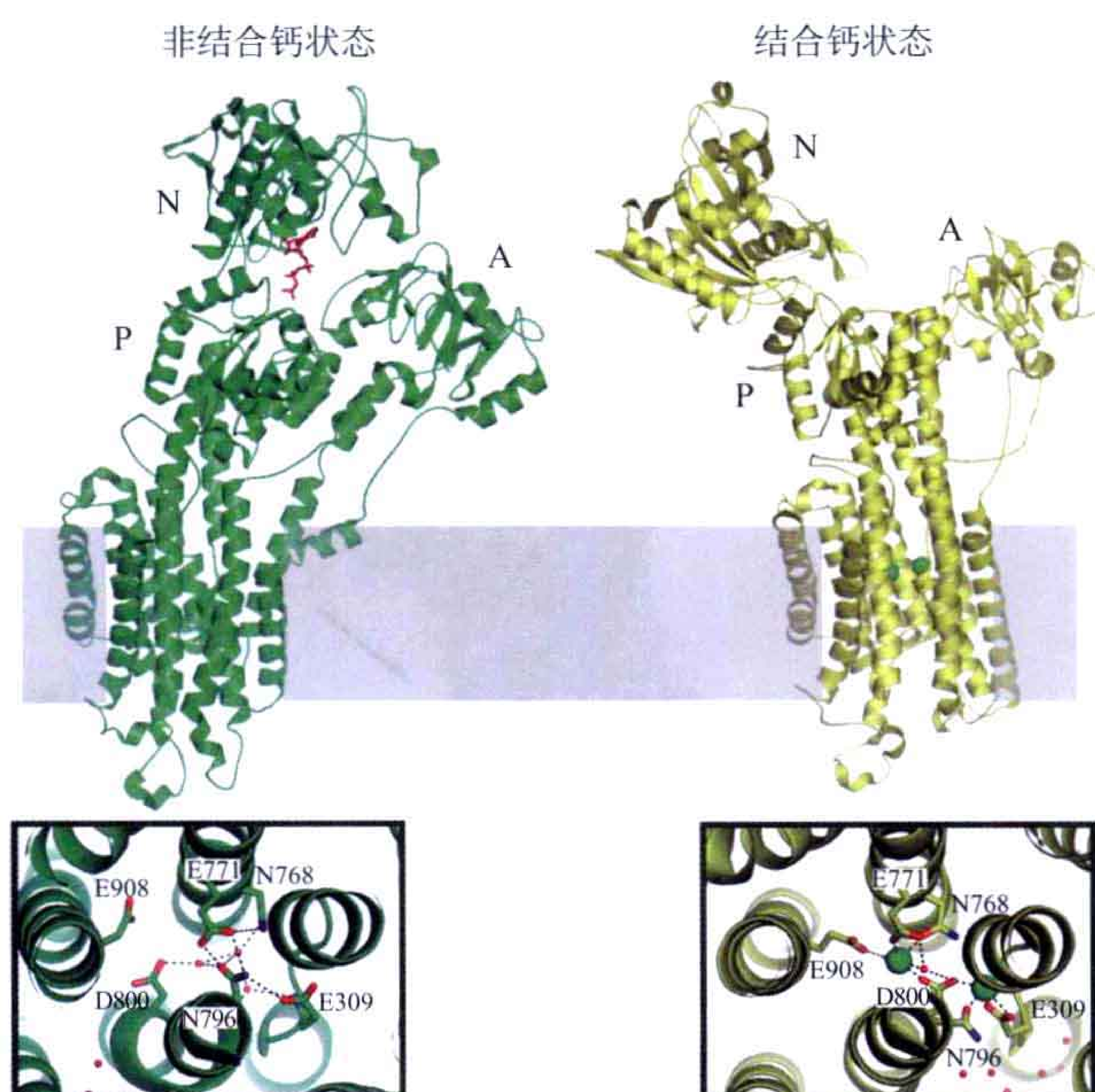


图 10.20　钙泵在不结合钙、质子化状态下的结构（左，PDB：1IWO）和钙结合状态的结构（右，PDB：1SU4）。结构下的方框里面显示离子结合位点的细节。这两个结构是把 C 端螺旋，即 M7 ~ M10 比对在一起来保证相同的视角，结构显示在钙结合后有很大的构象变化。此外，还有结构（PDB:2C88）解析在没有钙结合的状态下钙泵的非活性磷酸化位点结合一个 ATP 类似物（AMPPCP）。

ATP 结合状态及磷酸化过渡态（利用 ADP 和氟化铝模拟）下的结构展示了钙转运和磷酸化过程偶联的基础。钙离子进入结合 ATP 的蛋白质时激活磷酸化位点，而磷酸化过程偶联着钙离子进入通道的关闭（图 10.21）。相似地，去磷酸化过程则偶联着反向运输过程离子的包被 (通常称为 “occlusion”)，在钙泵中这个反向离子是质子。

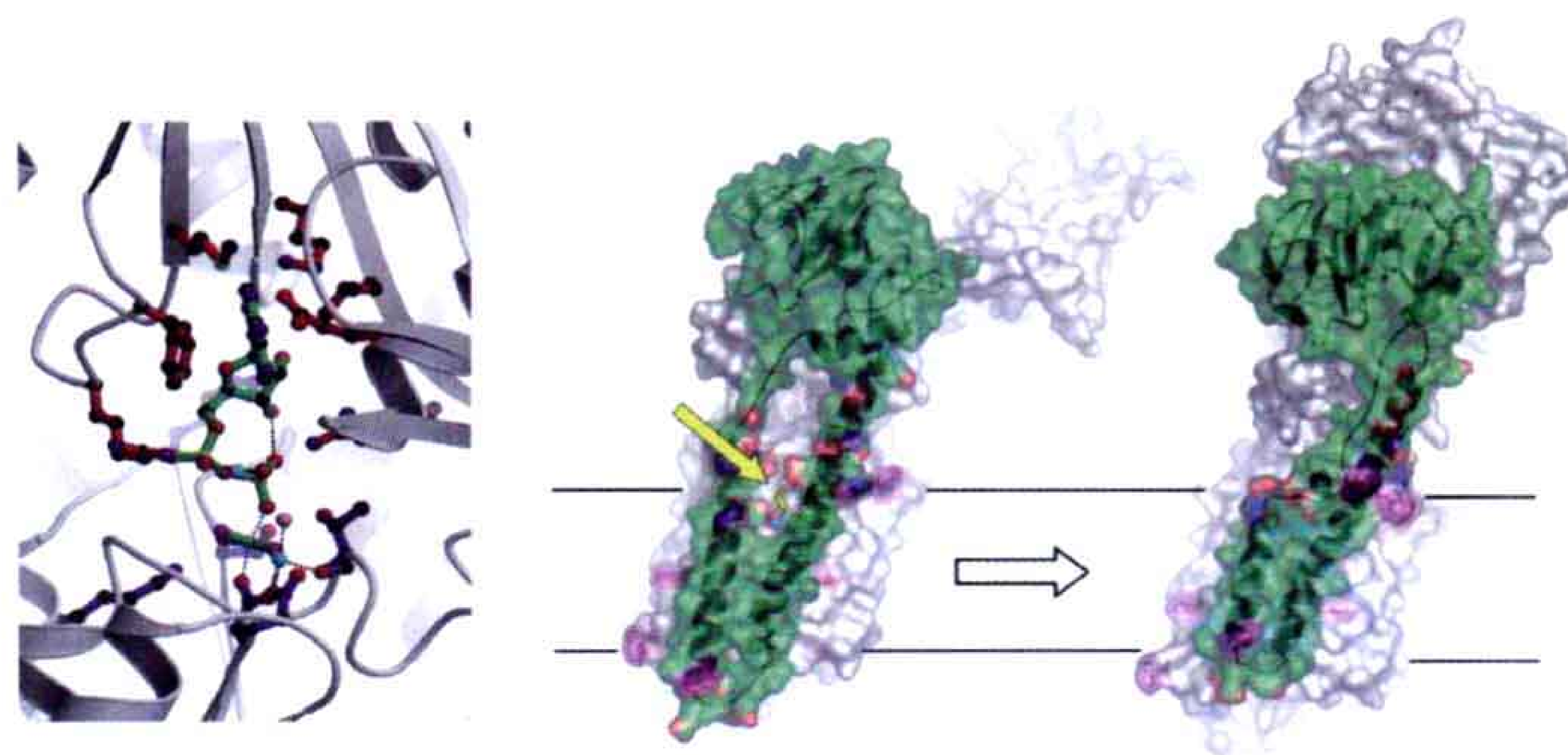


图 10.21　在钙泵里面 ATP 磷酸化和钙离子包埋的偶联关系。左：钙离子结合状态下钙泵与 ADP：AlF4⁻ 复合物（PDB：1T5T）展示的磷酸化过渡状态的细节图。N 结构域（红色）和 P 结构域（蓝色）的残基稳定住磷酸转移反应。保守的 Asp351 的磷酸化催化位点有 Asp351 的侧链和 ATP 的 γ 磷酸（这里是氟化铝）结合的一个镁离子（天蓝色小球），这样使 Asp351 的侧链受到磷原子的亲核攻击。另一个镁离子稳定着将释放的 ADP，这是在寡聚酶和核酸酶里面出现过的双镁离子催化的一个例子。右：显示由磷酸转移引起的胞质区向膜内钙结合位点通道（黄色箭头指示）的关闭（PDB：1SU4，1T5T）。这个关闭是由 M1 和 M2 这两个螺旋及 A 结构域的旋转调节的。而 A 结构域的转动则是在 N 结构域和 P 结构域间组装磷酸化位点的结果。

钙离子的释放发生在 E1P 到 E2P 磷酸化状态的改变过程中，这个变化打开一条释放通道，同时也打乱钙结合位点，使其与钙离子亲和力下降几个数量级。这个构象变化依赖于磷酸化中间态的形成，这也解释了转运过程如何可以定向并逆电化学梯度进行（图 10.22）。

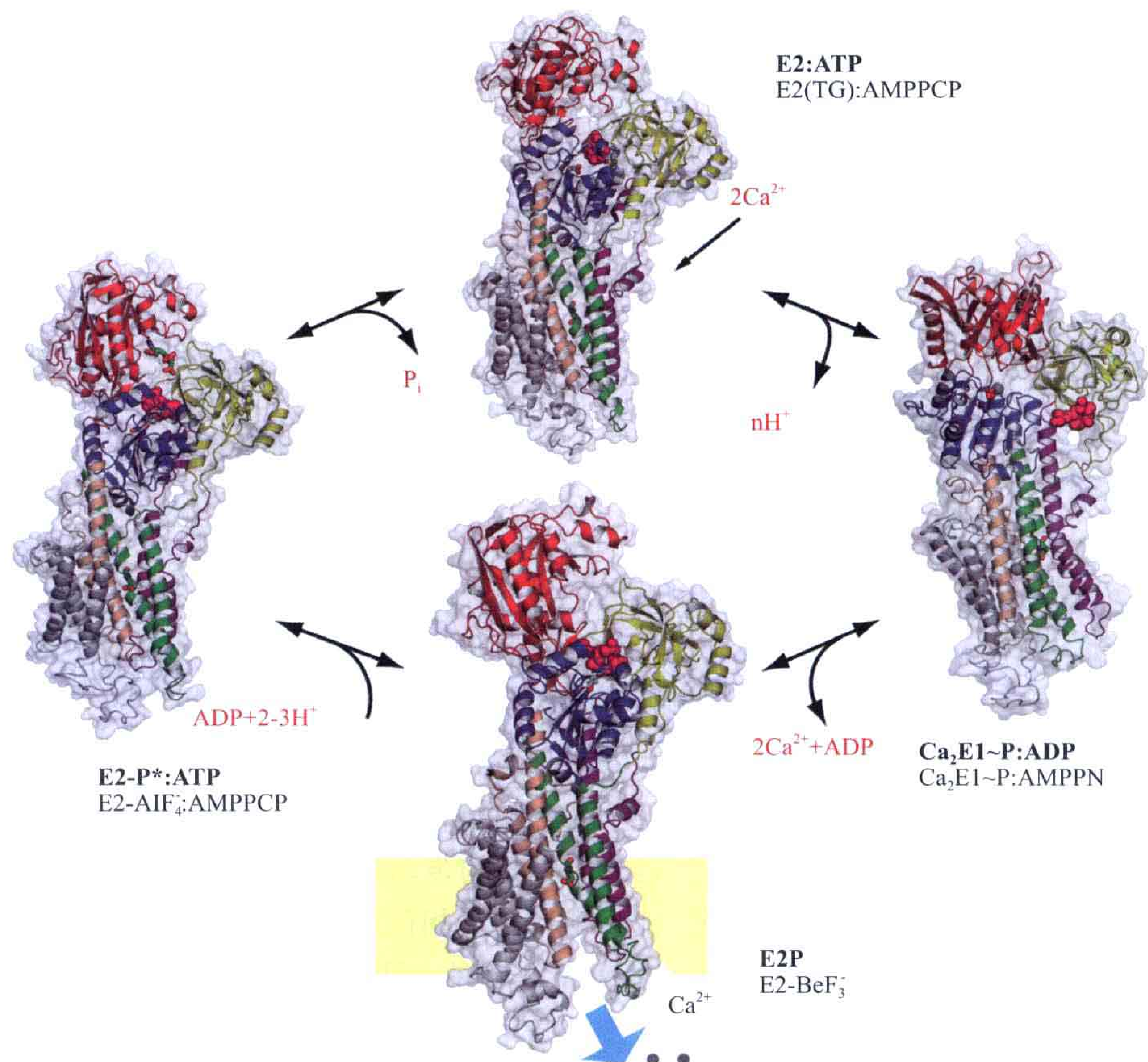


图 10.22　钙泵的工作循环。这个循环利用了结构被解析出来的 4 个状态。磷酸化过程和钙离子结合及包被过程相偶联，从 E1P 到 E2P 的异构化和钙离子释放过程相偶联，E2P 蛋白的去磷酸化和质子结合及包被过程相偶联，而 E2 到 E1 的异构化和质子释放过程相偶联。[原图来自 Olesen C, Picard M, Winther A-M, Gyrup C,Morth JP, Oxvig C, Møller JV, Nissen P. (2007) The structural basis of calcium transport by the calcium pump. *Nature* **450:** 1036-1042. Copyright (2007) Nature Publishing Group，在获得允许后做了修改。]

10.6　次级转运蛋白

调控 pH 和离子浓度就像摄取代谢物和营养物质一样对所有活细胞具有关键作用。次级转运蛋白是两个或者更多分子或离子的偶联转运，它们对于跨膜转运很关键。在任何情况下转移反应都应该遵守热力学规律。共转运蛋白朝同一方向转运不同的组分，

它们是共同转运体。通常一种离子（正电荷或者负电荷）会顺电化学梯度转运，使得另一种分子（离子）朝同一个方向但是逆电化学梯度转运。这类主动转运又被称为协助扩散。一个典型的例子是葡萄糖转运蛋白家族成员利用钠离子的顺势梯度来逆浓度转运葡萄糖。

此外还有一类膜蛋白家族用相反的方式行使功能，它们被称为反转运体，利用一种离子的顺势电化学梯度来使另一种分子（或者离子）朝相反的方向转运。一个典型的例子为钠钙交换蛋白（NCX）。在钠离子顺浓度梯度转运的同时，钙离子则是从另一个方向（反向转运）逆浓度梯度转向胞外。这是快速去除细胞内高浓度钙离子的一个重要的机制。在钠钙交换蛋白工作时，每从胞内转运出一个钙离子同时转运进来三个钠离子，这样还以 3 ： 2 交换正电荷的方式利用了跨质膜的负膜电位。

次级转运蛋白的离子结合位点位于膜中间，一些跨膜螺旋在这附近变得不连续，延伸出几个残基的肽段。短螺旋的双极性（见 2.1.4.3 节）为蛋白质中间被转运离子结合位置提供部分带电性。在疏水核心的中间，没有多少机会进行极性相互作用。延伸的肽链为离子与主链氨基或者羧基结合提供可行性。此外，离子结合位点也在疏水中心有一些极性的侧链。不同位点的离子选择性很高，取决于离子半径和不同离子对配体的倾向性。选择性很大程度上源于结合位点对被选择离子的吻合度。

有趣的是一些膜通道和次级转运蛋白的结构呈现一种反式二重对称性，二重轴平行于膜，提示基因重复的发生是转运蛋白形成的基础。对称部分间的弱序列相似性也证明了这个观察，由反式对称单体形成的二聚体能为形成一个向内和向外的构象提供结构基础。结合电化学梯度，特定转运底物也许在一面结合较强而在另一面结合较弱，这样就构成了一个次级转运蛋白。

10.6.1　神经递质：钠离子共转运体——NSS 家族蛋白

有一类共转运体家族能利用跨膜钠离子梯度向细胞内转运氨基酸。在动物中，这类家族的转运体进一步负责从突触间隙中回收神经递质，这类家族因此被称为神经递质：钠离子共转运蛋白（NSS）家族。NSS 家族膜蛋白转运的分子有儿茶酚胺类如多巴胺、羟色胺、去甲肾上腺素和 γ 氨基丁酸（GABA）。神经递质转运蛋白的机能失常会导致一些如癫痫、抑郁症和自闭症等疾病。利用神经药物抑制这类转运蛋白能延长神经递质在突触间隙中的活性。例如，羟色胺转运体能被“特异性羟色胺回收抑制剂”（SSRI 药物）来抑制，这些药物包括氟西汀（百忧解）和西酞普兰（喜普妙）。三环抗抑郁药（TCA）如氯米帕明是另一类具有重要临床意义的抑制剂家族，但是它们特异性较低。在许多细菌和古细菌中也发现了与人类 NSS 家族蛋白有显著序列同源性的氨基酸转运蛋白，其中一些也能被三环抗抑郁药抑制。

10.6.1.1　LeuT

来自 *Aquifex aeolicus* 菌的 LeuT 参与把疏水氨基酸转运到胞内的过程。结构显示有

12 个跨膜螺旋以二重反式拓扑排列。跨膜螺旋 1~5 和螺旋 6~10 之间有二重对称性，对称轴在膜里。在膜两面的环里存在一些额外的螺旋。在膜中间核心螺旋是不连续的。底物亮氨酸和两个钠离子的结合位点位于蛋白质中心，在螺旋 1 和 6 的延伸部分（图 10.23）。只有在由两个螺旋组成的胞外环发生结构变化的情况下，亮氨酸才能从结合位点被转运出去。这个环的外面是一个腔室，这个腔室被认为是从胞外环境到结合位点的进入通路的一部分。在对称性上腔室内部残基的对应残基是被包被起来的，但是其可以在另一个构象下的腔室里面暴露出来。神经递质的 TCA 抑制剂能够和 LeuT 结合并且抑制亮氨酸转运。不过，抑制剂并不是去取代结合的亮氨酸而是结合在进入通道上，稳定住封闭的构象，这样就中断了工作循环。分析另外两种三环抗抑郁药（米帕明和去甲丙咪嗪）的结构发现它们采用了同一个结合位点。

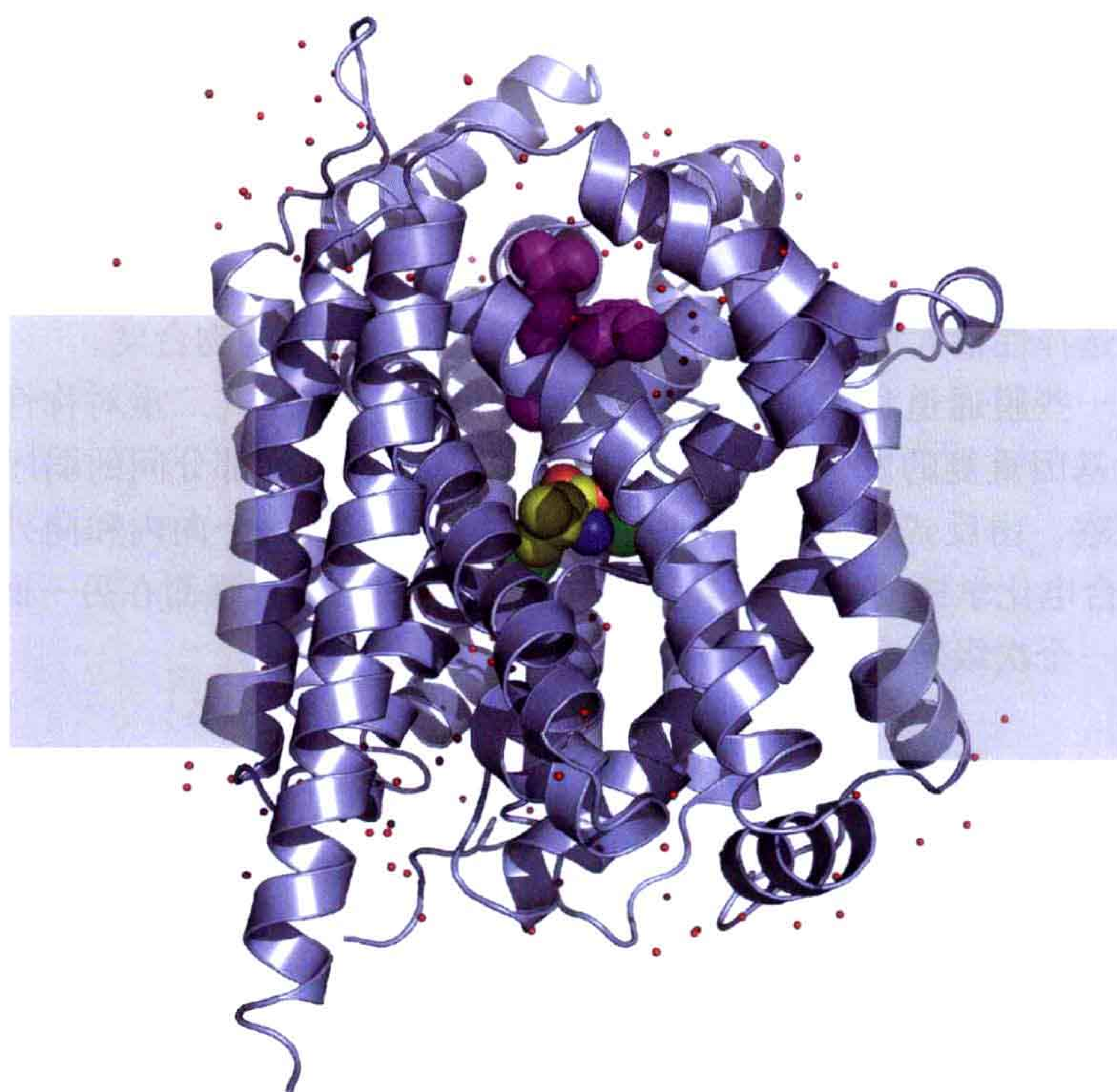


图 10.23 封闭状态下的 LeuT 结构（浅蓝色）。与两个钠离子（绿色小球）作用的亮氨酸底物（黄色、红色和蓝色小球）结合在膜中间的封闭位点上。氯米帕明抑制剂用红色小球表示。它堵住了胞外侧的进入通道，稳定了蛋白质的封闭状态（PDB：2Q6H）。

LeuT 及 NSS 家族的其他成员至少要有三个状态：向外开口、封闭和向内开口（图 10.24）。胞外侧和胞内侧的门调控与两侧的接触。胞外侧的门仅由几个氨基酸残基组成，而在封闭状态下与胞内侧的接触被约 25Å 的规则蛋白结构阻挡着。三环抑制剂封堵了胞外侧通道。特别是它们稳定了 Arg30 和 Asp404 间的盐桥。这些残基在一定程度上同时与底物及一个钠离子相互作用。目前没有在实验上研究出这些不同的结构状态，不过那些不连续的螺旋可能是结构变化的一部分。

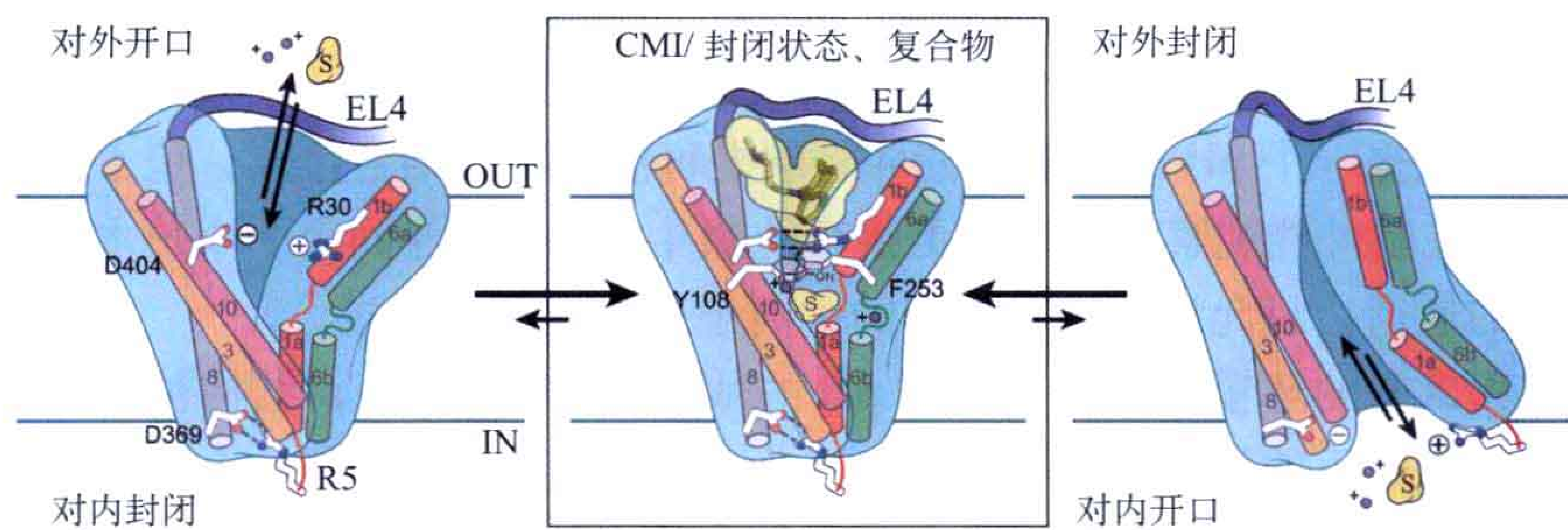


图 10.24　LeuT 共转运体转运亮氨酸和两个钠离子的一个可能的机制。至少需要三种状态：向外开口时可以与胞外溶液交换亮氨酸和钠离子；封闭状态下被转运离子在 LeuT 里被包被起来；向内开口时可以与胞内溶液交换亮氨酸和钠离子。三环抑制剂把转运蛋白锁在封闭状态 [原图来自 Singh, SK, Yamashita, A, Gouaux, E. (2007) Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448:** 952-956. Copyright (2007) Nature Publishing group，在获得允许的情况下略有修改]。

10.6.2　反转运蛋白

许多反转运蛋白参与跨胞质膜的离子交换。在细菌里面，一类反转运蛋白以交换质子的方式排出钠离子或者锂离子。质子的梯度是构建钠离子跨膜电化学电势的驱动力。反过来，钠离子梯度又被用于其他如鞭毛转动的过程。反转运体的第一个结构来自一个叫 NhaA 的大肠杆菌蛋白。

10.6.2.1　NhaA 反转运蛋白

NhaA 是一个钠离子 / 质子反转运蛋白。它被 pH 所调控，在 pH 6.5~8.5 时被激活。NhaA 也是一个由 12 个跨膜螺旋组成的二聚体（图 10.25）,N 端和 C 端都在细胞内。它们以不同于 NSS 家族共转运体的独特的方式排列。不过分子的一部分在与膜平行方向有二重对称性。这包括螺旋Ⅲ、Ⅳ和Ⅴ，它们与螺旋Ⅹ、Ⅺ和Ⅻ对称性相关。

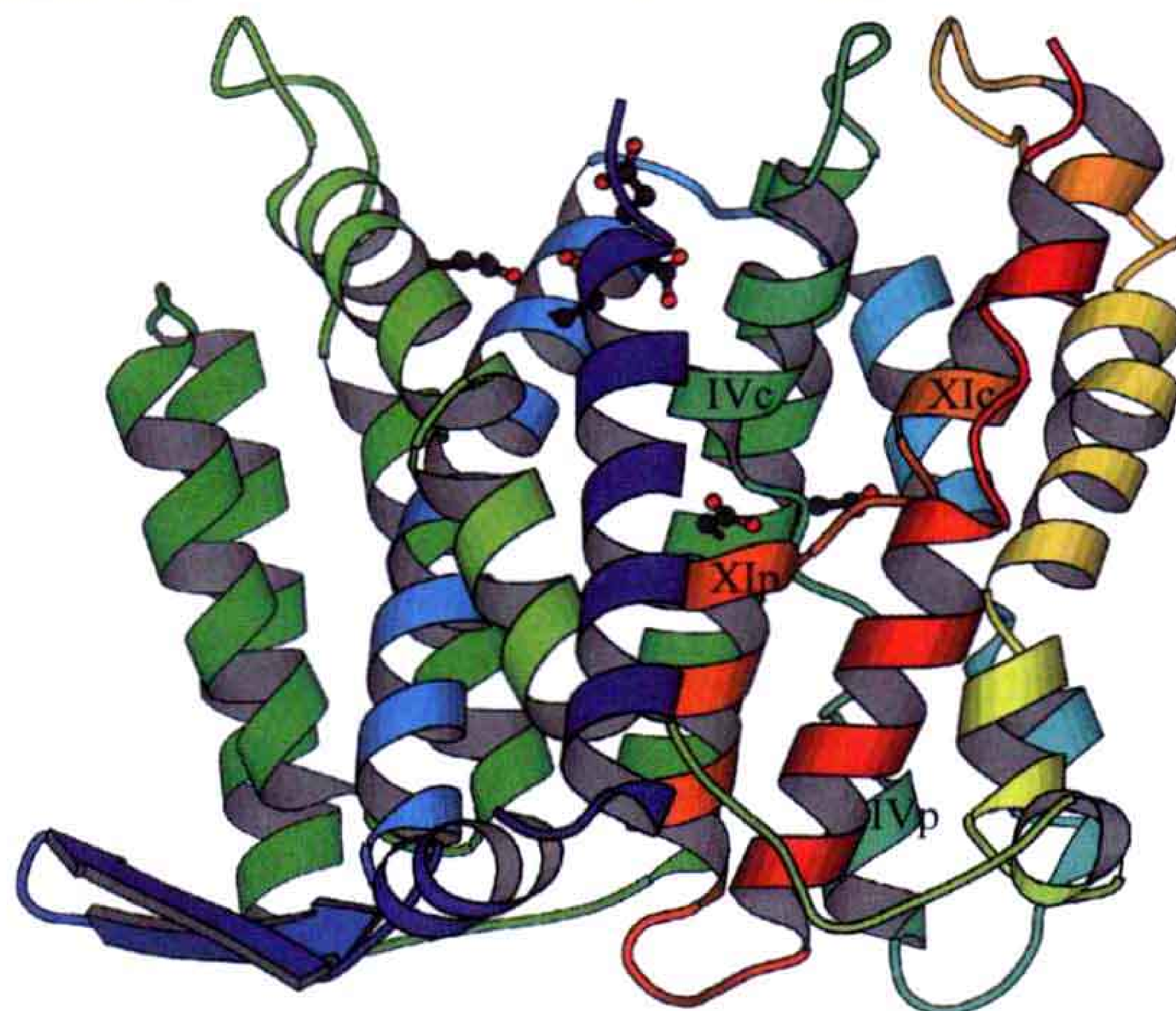


图 10.25　大肠杆菌反转运蛋白 NhaA 的结构（PDB：1ZCD）。

螺旋Ⅳ和Ⅸ是不连续的，有部分是与延伸的肽链相连的。这些肽链片段大概在膜中间相互交叉。在胞内侧有一个漏斗形离子通道，其上排列着一个负电残基群。这个漏斗形结构可能是一个 pH 感受器。漏斗一端是可能的离子结合位点，位于螺旋Ⅳ和Ⅸ的交叉部分。在胞外侧观察到了一个更狭窄的漏斗。两个漏斗间由一个疏水屏障隔开，有一条狭窄的通道，只允许非水合状态的钠离子或者锂离子通过。

螺旋Ⅷ和Ⅸ间的环里面有参与感受 pH 的残基。在低 pH 下，结构被锁定了。目前只有这种锁定状态下的结构被解析出来。在生理 pH 下这个环参与这个蛋白质的多重功能：参与二聚体的形成，参与形成指向离子结合位点的胞内漏斗，还决定质子 / 钠离子的转运量。这个环上残基的突变影响蛋白质对 pH 的依赖性，并改变其对钠离子的结合强度。

延伸阅读

原始文献

Deisenhofer J, Epp O, Miki K, Huber R, Michel H. (1984) X-ray structure analysis of a membrane protein complex.*J Mol Biol* **180**: 385-398.

Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**: 69-77.

Ferguson AD, Chakraborty R, Smith BS, Esser L, van der Helm D, Deisenhofer J. (2002) Structural basis of gating by the outer membrane transporter FecA. *Science* **259**: 1715-1719.

Henderson R, Unwin, PNT. (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**: 28-32.

Jiang Y, Lee A, Chen J, Cadene M, Chait BT, MacKinnon R (2002) Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**: 515-522.

Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauss N. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. *Nature* **411**: 909-917.

Long SB, Campbell EB, Mackinnon R. (2005) Crystal structure of a mammalian voltage-dependent Shaker family K+channel. *Science* **309**: 897-903.

Olesen C, Picard M, Winther AM, Gyrup C, Morth, JP, Oxvig C, Mϕller JV, Nissen P. (2007) The structural basis of calcium transport by the calcium pump. *Nature* **450**: 1036-1042.

Toyoshima C, Nakasako M, Nomura H, Ogawa H. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. *Nature* **405**: 647-655.

Walz T, Hirai T, Murata K, Heymann JB, Mitsuoka K, Fujiyoshi Y, Smith BL, Agre P, Engel A. (1997) The threedimensional structure of aquaporin-1. *Nature* **387**: 624-627.

Yue WW, Grizot S, Buchanan SK. (2003) Structural evidence for iron-free citrate and ferric citrate binding to the TonBdependent outer membrane transporter *FecA*. *J.Mol.Biol*. **332**: 353-368.

综述文章

Elofsson A, von Heijne G. (2007) Membrane protein structure: prediction versus reality. *Annu Rev Biochem*. **76**: 125-140

Gouaux E, Mackinnon R. (2005) Principles of selective ion transport in channels and pumps. *Science* **310**: 1461-1465.

Møller JV, Nissen P, Sørensen TL, le Maire M. (2005) Transport mechanism of the sarcoplasmic reticulum Ca2+ -ATPase pump. *Curr Opin Struct Biol*. **15**: 387-393.

Iwata S, Barber J. (2004) Structure of photosystem II and molecular architecture of the oxygen-evolving centre. *Curr Opin Struct Biol*. **14**: 447-453.

数据库链接

Membrane proteins of known structure (Stephen White's laboratory, Irvine, California, USA): http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html

Membrane protein data bank (M. Caffrey and coworkers, Limerick, Ireland):
http://www.mpdb.ul.ie/

Protein Data Bank of Transmembrane Proteins (Budapest, Hungary):
http://pdbtm.enzim.hu

（刘翔宇　译校）

第十一章 信号转导

11.1 信号转导：从外部控制细胞活性

真核细胞的诸多生理活动必须受到细胞外各种因子的调控。对于多细胞生物来说，这种胞外调控对生理过程显得十分重要，因为细胞需要分化生成具有特化细胞的各种组织，而处于特定组织中的细胞通过响应环境信号来帮助整个组织发挥正常功能。

胞外调节是通过信号通路（signal pathway）来影响细胞功能的。在大多数信号通路中，胞外因子结合在细胞膜上的受体分子上。受体分子将信号而非分子本身传递到胞内。在细胞内，众多其他分子相互作用形成一个信号通路，感受信号并且放大信号效应，最终完成所需的细胞活动。其中，胞外因子可以是小分子（如肾上腺素、气味分子等），也可以是短肽（如胰高血糖素），或者是蛋白质（如生长激素、干扰素等）。在其他一些信号通路中，脂溶性分子也可作为胞外因子，这些脂溶性分子可以穿过细胞膜，激活胞内受体。上述整个过程就称为信号转导。

信号通路可分为以下两类。

（i）一些信号需要细胞针对刺激做出状态的快速变化，如人眼视觉细胞对光子的反应。在这个例子中，信号导致了视觉细胞状态的快速变化，这一变化被神经细胞感受并转化成脉冲传递到大脑。信号的转导通常导致一些蛋白质活性的变化。这一类信号通路通常利用 G 蛋白偶联受体（GPCR）和可以激活多种效应分子的 G 蛋白三聚体。

（ii）另一类信号导致细胞发生更加持续的性质变化，并改变特定蛋白的胞内含量。这类信号通路可以使细胞核内转录阻抑蛋白的激活或失活。在这一类通路中，胞外信号通常是其他细胞分泌的激素或生长因子。

在细胞内部，蛋白质的磷酸化和去磷酸化（见附录 E）是控制细胞生理功能的两种主要方式。许多信号通路中都包含蛋白激酶，它们可以磷酸化酪氨酸或者丝氨酸 / 苏氨酸侧链羟基。这些激酶反之也可以通过自身含有的酪氨酸、丝氨酸 / 苏氨酸的磷酸化和去磷酸化而受到调控。特定的磷酸酶可以水解掉被磷酸化的激酶上的磷酸基团，从而反转该激酶的活性，而这些磷酸酶同时又可能受到胞外信号的调控。

许多信号通路都是利用 G 蛋白的特点来发挥功能的。G 蛋白在细胞内发挥分子闸的作用。它们之所以存在于众多信号通路中，是由于它们可以被开启而处于活性（开）状态。而这一过程又受到该通路中其他受体和蛋白质的调节，使 G 蛋白回到“关闭”状态。

至少有两类重要的 G 蛋白参与信号转导过程，一类是与 G 蛋白偶联受体相关联的三聚体 G 蛋白；另一大类是单体 G 蛋白——RAS 超家族（表 11.1）。

表 11.1　RAS 超家族中的单体 G 蛋白及其功能举例

Ras	控制基因表达的胞内信号通路
Rho	调控细胞骨架的生长
Rab	调控细胞内囊泡的运输
Ran	调节细胞入核及出核过程
Arf	调控囊泡运输

信号是怎样穿过细胞膜进行通讯传导的呢？所有的细胞膜受体需要将与胞外配体的结合过程转化成膜内的信号。 细胞膜外侧的变化会影响到胞内蛋白结构域的构象变化或重排。这一过程可以通过如下方式进行：

- 受体的二聚化或二聚体接触面的变化。
- 单体受体蛋白的构象改变。
- 通道。

为了更好地解释信号通路的工作机理，我们将通过几个具体的例子来进行说明：细胞因子受体通路、酪氨酸激酶受体通路、G 蛋白偶联受体通路。

11.2 细胞因子介导的信号传递

细胞因子受体（cytokine receptor）作为受体分子可以结合如干扰素 (interferon)、白细胞介素 (interleukin) 和激素 (hormone) 等不同的小分子质量蛋白，并诱导细胞的分化、分裂及凋亡。受体分子通常是膜蛋白，由胞外组分、单次跨膜部分和膜内结构域组成。有些受体分子，如激素受体，是由两个相同的蛋白质分子构成的；而有些受体分子则是由两个不同的受体分子结合两个底物分子而组成的一个活性复合体，这两个不同的受体分子通常分别被称为 α 亚基和 β 亚基。受体的胞外部分含有一个受体结合结构域，根据不同受体家族的特点，胞外部分也会存在其他的结构域。

细胞因子与受体结合并诱导受体分子单体的二聚化是细胞因子受体通常的激活方式。 受体的激活导致了其胞内结构域被类 Janus 激酶家族中的酪氨酸激酶 JAK1 和 JAK2 磷酸化。 然后 STAT （signal transducer and activator of transcription，信号转导和转录激活因子）分子被募集到磷酸化的受体分子上，并通过后续的步骤调控细胞核内的基因转录。这就是称为 JAK-STAT 的信号通路。

11.2.1　生长激素受体介导的信号转导

生长激素（GH）进行的信号转导是一个较为简单的例子。一种类型的受体分子在激素结合后发生二聚化，以酪氨酸激酶 JAK2 磷酸化受体为开端，导致一系列的磷酸化级联反应（图 11.1）。

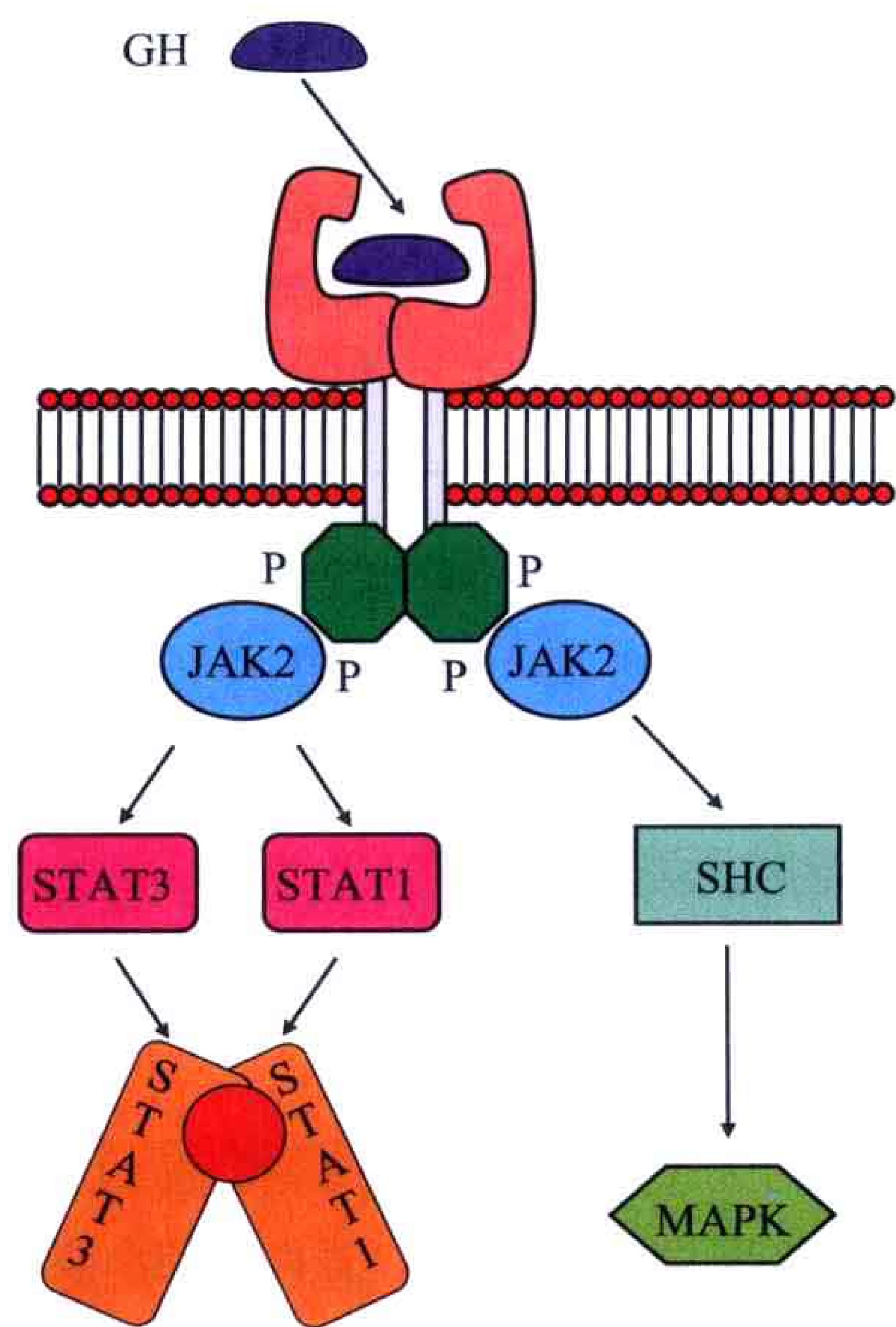


图 11.1　生长激素信号通路概况。生长激素（GH）的结合导致了受体分子的二聚化。随后，酪氨酸激酶 JAK2 结合到受体二聚体上，并磷酸化受体胞质部分。其他的蛋白质随后结合到受体上并被 JAK2 磷酸化。STAT 家族蛋白被磷酸化后可很快二聚化，从而进入细胞核调控多种基因的转录（图中红色所示为 DNA 结合区）。酪氨酸激酶 SHC 被活化，通过几步后续反应激活 MAP 激酶，MAP 激酶又可以活化其他更多转录因子。

生长激素受体的胞外部分与其配体生长激素的复合体晶体结构已经得到解析。激素受体的胞外部分是由两个结构域（D1 和 D2）组成的，这两个结构域都有Ⅲ型纤连蛋白折叠模式 (fibronectin type Ⅲ fold) （图 11.2）。这类折叠模式在受体分子的胞外部分很常见（表 11.2），它与 Ig- 折叠（Ig-fold，免疫球蛋白折叠）（见第十四章）很相似。

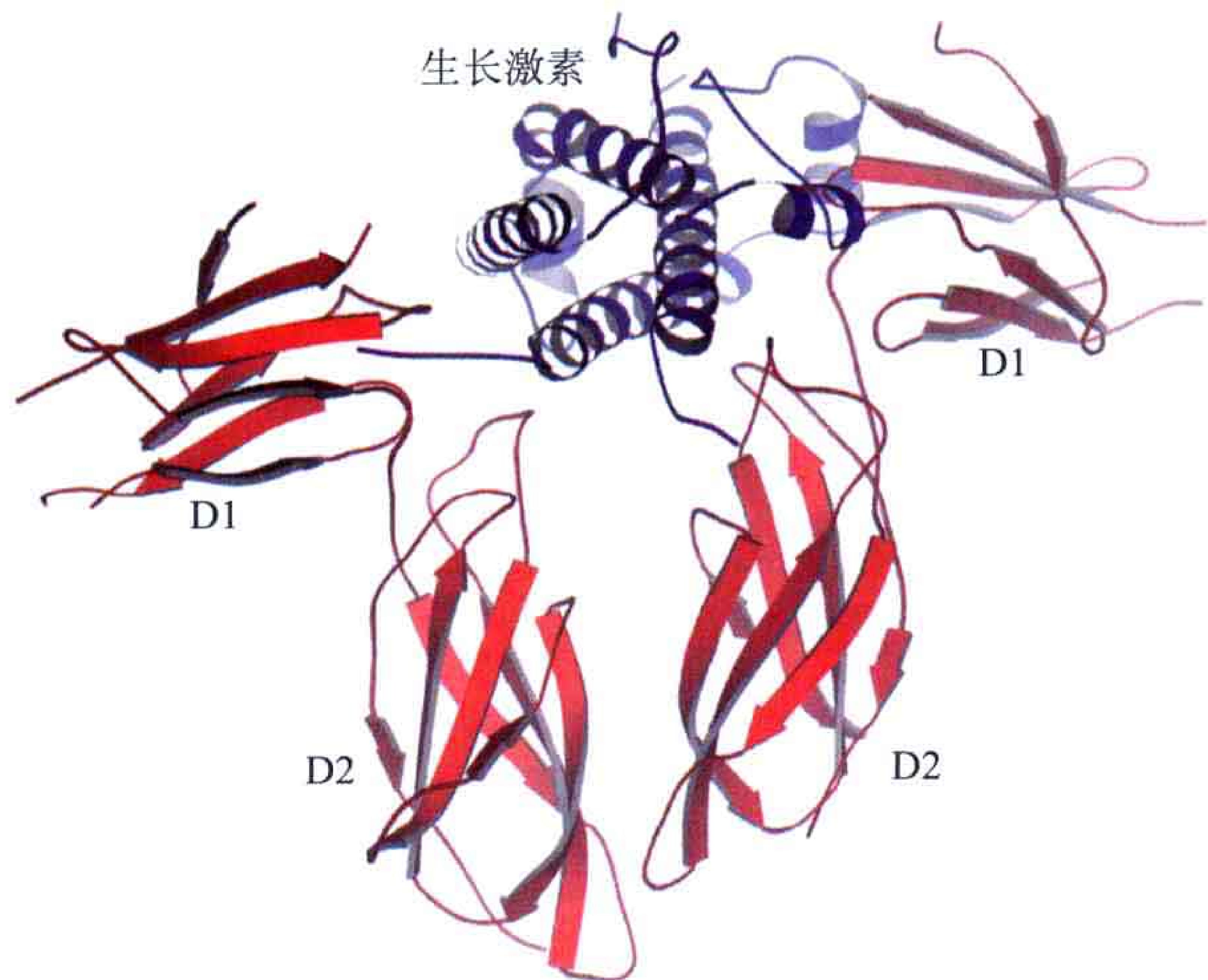


图 11.2　激素受体胞外区和与之相结合的生长激素。生长激素信号是由受体分子的二聚化形成的。激素分子在三维结构上是 α 螺旋束（α helix bundle），许多其他细胞因子也含有该结构。受体分子两个相同单体分子中都含有两个类似结构域（D1、D2），该结构域具有Ⅲ型纤连蛋白折叠模式。跨膜螺旋区与胞内部分的连接处位于图的底端（PDB：3HHR）。

表 11.2　受体分子胞外区域蛋白结构域与其他胞外蛋白举例

结构域名称	大小和构象	含有该结构域的蛋白质举例
Fn1 (fibronectin type Ⅰ，Ⅰ型纤连蛋白)	约 40 个残基，由二硫键连接的两个小的片层结构	纤连蛋白等其他胞外结合蛋白
Fn2 (fibronectin type Ⅱ，Ⅱ型纤连蛋白)	约 40 个残基，由二硫键连接的两个小的片层结构	纤连蛋白，存在于血液凝集物中的蛋白质
Fn3 (fibronectin type Ⅲ，Ⅲ型纤连蛋白)	约 100 个残基，反平行 β 三明治结构	纤连蛋白，细胞因子受体，酪氨酸激酶受体和其他细胞膜表面结合蛋白
receptor L domain	约 190 个残基，β 螺旋	受体酪氨酸激酶
C（furin-like cysteine-rich domain，类弗林半胱氨酸富集结构域）	约 120 个残基，几个小的 β 发夹，由二硫键稳定整体构象	受体酪氨酸激酶
Ig (immunoglobulin，免疫球蛋白)	约 100 个残基，反向平行 β 三明治结构，与 Fn3 十分相似	受体酪氨酸激酶，多种免疫系统中的蛋白质分子

从结构上可以看到一个激素分子结合到两个受体分子上，形成 1 ：2 的复合体模式。来自于每个受体分子结构域的一端卷曲（loop）区形成底物结合表面。结构域 D1 的底物结合卷曲区对应于免疫球蛋白结构中的高变区，而 D2 结构域将 β 三明治（β sandwich）结构的另一端暴露于激素配体。

配体的结合产生受体分子二聚化的信号。生长激素分子是不对称的，但是受体分子与配体结合的元素是相同的。因此配体分子的不对称性就造成了配体受体复合物中受体分子二聚体的不对称性。我们知道两个大分子的结合亲和力的大小基本与其发生结合的表面面积成正比。在激素与其受体结合复合物中，一个受体分子与激素的作用面要大于另一个分子。这样就可以推断配体介导的受体分子二聚化的机制：在没有配体存在的情况下，受体分子单体之间的结合亲和力比较小。在激素存在时，一个受体单体分子与配体激素高亲和力结合。此时，复合物已经暴露出足够大的结合面，该结合面包括了激素分子和两个受体单体分子的结合元素，这个结合面保证了受体的另一个单体分子与配体的高亲和力结合。只有该配体－受体复合体形成后，受体分子的胞内部分才会具有与 JAK2 分子相互作用的正确取向，此时 JAK2 被激活并磷酸化受体胞内部分。生长激素受体和其他细胞因子受体的胞内部分约有 350 个残基，其三维结构仍然未知。

配体结合产生的信号在 STAT 分子上得到呈递，STAT 分子属于多结构域转录因子中的一个家族。STAT 分子含有 SH2 结构域（SH2 domain），该结构域可以特异地结合在含有磷酸化酪氨酸的一段肽段上（图 11.3）。因此 STAT 分子就可以结合在含有磷酸化酪氨酸的受体分子上。STAT 分子在靠近 SH2 结构域的尾部含有一个酪氨酸，该残基可以被 JAK2 激酶磷酸化。然后 STAT 通过 SH2 结构域二聚化，并穿过核膜进入细胞核。在此处，利用其含有的 DNA 结合结构域 (DNA-binding domain) 与 DNA 上特定的位点结合并激活一些基因的转录 (图 11.3)。

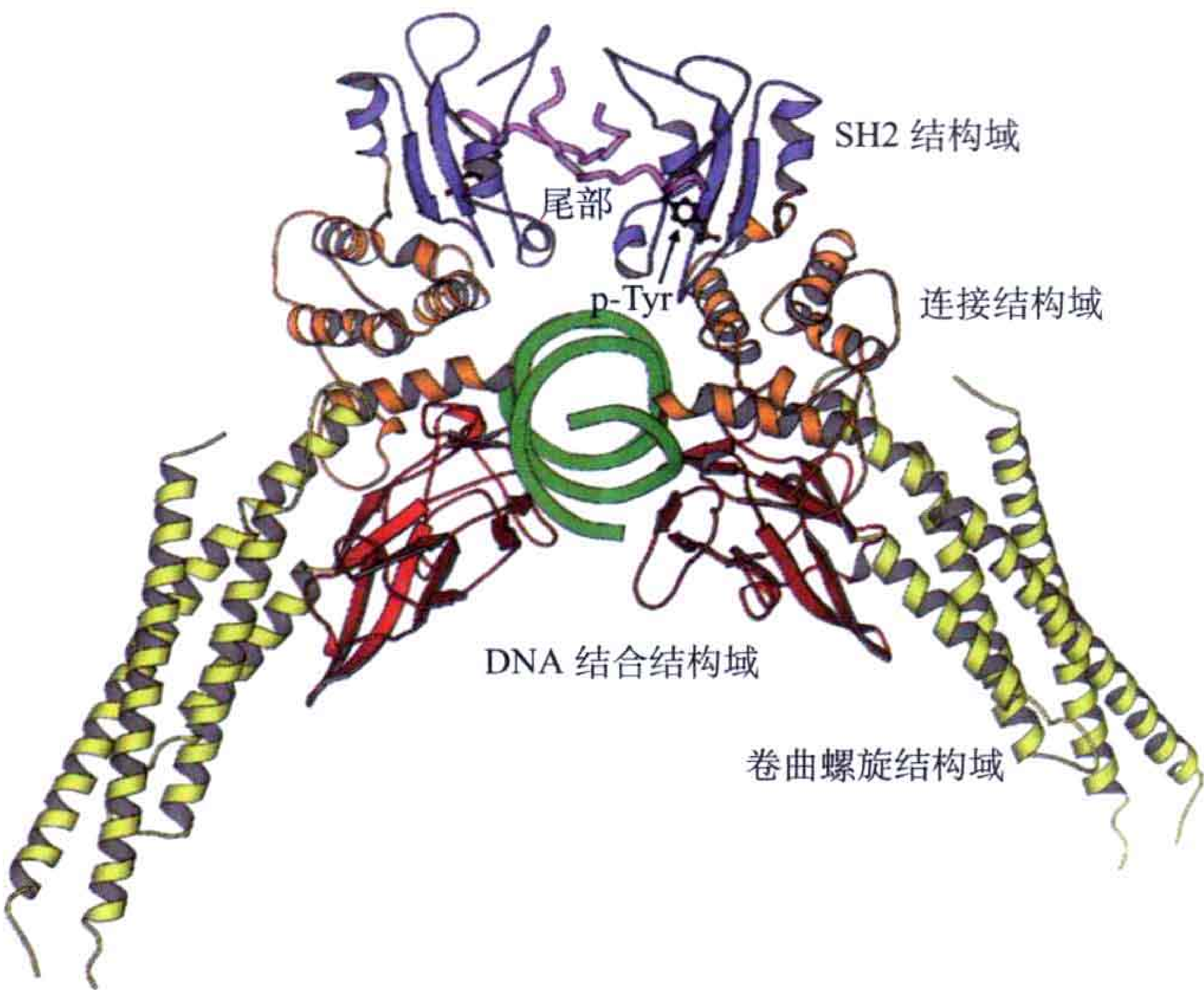


图 11.3 胞内活性：STAT3 二聚体与一段双链 DNA（绿色）相结合的复合物。该图中，二重轴的方向差 10° 与纸面垂直。磷酸化的酪氨酸（p-Tyr）介导了 STAT3 的二聚化。该残基位于蛋白质的羧基端尾部（淡紫色），并与另一个分子的 SH2 结构域 (SH2 domain)（蓝色）相结合。连接 SH2 结构域与尾巴间的氨基酸在结构上处于无序状态。DNA 结合结构域 (DNA-binding domain)(红色) 是与 p53 转录因子相似的免疫球蛋白类折叠。它位于卷曲螺旋(coiled coil)结构域(黄色)的前面，其后接有一段连接结构域(橙色)。在该模式图中，氨基端结构域没有包括在其中(PDB:1BG1)。

11.2.2 酪氨酸激酶活性的调控

参与信号通路的蛋白激酶可以磷酸化酪氨酸或者丝氨酸 / 苏氨酸残基。这类激酶属于酶类的一个大家族。激酶蛋白通常含有两个结构域：一个结构域含有由 5 个反向平行 β 股构成的折叠片和一个 α 螺旋；另一个结构域包含 7 个螺旋（图 11.4）。活性中心位于两个结构域之间的缝隙中。

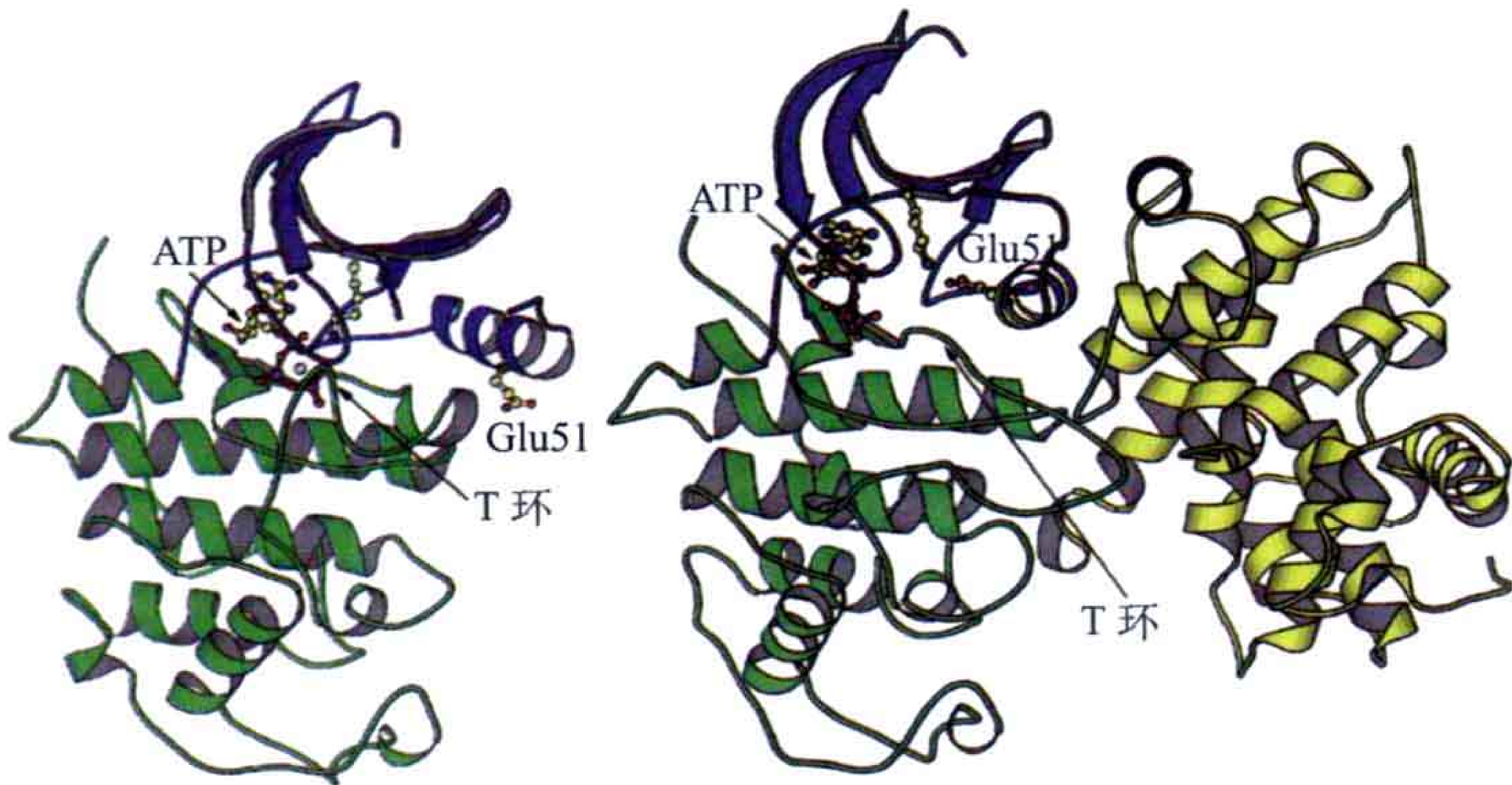


图 11.4 细胞周期蛋白依赖性激酶的失活和活性状态。激酶的两个结构域分别以蓝色和绿色标识。在失活的状态下（左），活性 T 卷曲（T loop）阻挡了活性中心使其不能与底物相结合。保守的第 51 位谷氨酸（Glu51）处于指向远离活性中心的位置。在活性状态中（右），在细胞周期蛋白（cyclin）(黄色) 的结合下，活性 T 卷曲从活性中心移开，位于上部的结构域中的螺旋发生旋转并移入活性中心。第 51 位谷氨酸的侧链与同底物三磷酸腺苷（ATP）十分接近的赖氨酸相互作用（PDB:1HCK 及 1FIN）。

酪氨酸激酶的调控发生在一段结构柔性的卷曲 (flexible loop) 上。当其阻挡活性中心时，激酶处于失活状态。这些卷曲被称为活性卷曲（activation loop）。目前，细胞周期蛋白依赖性激酶（cyclin dependent kinase，CDK）(图 11.4) 的激活机制已经阐明。在失活状态下，氨基端结构域中的一段螺旋远离活性中心；而在活性状态下，阻挡在活性中心的活性卷曲挪开，上述螺旋挪至活性中心。

11.2.3 利用结合结构域调节蛋白质活性

许多胞内蛋白分子是模块化的，即具有多个结构域。这些结构域的功能通常是将这些蛋白质带到细胞中有需要的地方，即带到具有特定酶活的蛋白质结构域靶标上，细胞膜上，或者其他特定的位置（表 11.3）。

表 11.3 胞内蛋白结合结构域举例

结构域名称	结构	结合特异性	含在信号通路中有该结构域的蛋白质举例
SH2（src homology domain 2, src 同源蛋白结构域 2）	约 100 个残基，反向平行片层加两边各一个螺旋（图 11.2）	含有磷酸化酪氨酸的肽段	γ 型磷脂酶 C，酪氨酸激酶（src），磷酯酰肌醇 -3- 激酶，酪氨酸蛋白水解酶 syp, RasGAP, GRB2
SH3（scr homology domain 3, src 同源蛋白结构域 3）	约 50 个残基，小的 β 桶（图 11.3）	富含脯氨酸肽段	磷脂酶 C，酪氨酸激酶（src），细胞骨架蛋白，RasGAP, GRB2
PH (pleckstrin homology) 结构域	约 120 个残基，β 三明治结构加羧基端螺旋	膜粘连及其他	磷脂酶 C，蛋白激酶，GEF 蛋白（Sos），GAP 蛋白，细胞骨架蛋白
PERM 结构域	约 300 个残基，3 个亚结构域	膜粘连	酪氨酸激酶，磷酸酶，细胞骨架蛋白（4.1, ezrin, moesin、radixin）
PTB(phosphotyrosine binding domain, 磷酸化酪氨酸结合结构域)	类似于 PH 结构域	含有磷酸化酪氨酸的肽段	Shc 接头蛋白，IRS-1 胰岛素受体底物分子
C2（protein kinase C domain, 蛋白激酶 C 结构域）	约 120 个残基，免疫球蛋白类折叠	膜粘连	磷脂酶 C，蛋白激酶 C，其他激酶，RasGAP
C1(protein kinase C domain, 蛋白激酶 C 结构域)	约 50 个残基，结合两个锌离子的富含半胱氨酸的蛋白质结构域	二酰油，佛波醇酯	蛋白激酶 C，其他激酶
PDZ（DHR 结构域，GLGF 重复）	80 个残基，β 桶	常以缬氨酸结束的短肽	信号转导系统中的大量蛋白质，通常串联存在
WD40 结构域	7 瓣螺旋桨		三聚体 G 蛋白 β 亚基，Arp2/3 复合物
EF hand（EF 手结构域）	约 100 个残基，形成 4 个 EF 手模体（EF-hand motif）	钙离子	磷脂酶 C

11.2.3.1 SH2 及 SH3 结构域

SH2 和 SH3 结构域在很多蛋白质中都存在。这个结构域是用非受体类酪氨酸激酶（non-receptor tyrosine kinase）scr（scr 同源蛋白）来命名的。src 和其他类似的激酶可以与细胞膜结合并可以激活众多细胞通路。它们通常含有一个蛋白激酶结构域、一个

SH2 结构域、一个 SH3 结构域（图 11.5）。

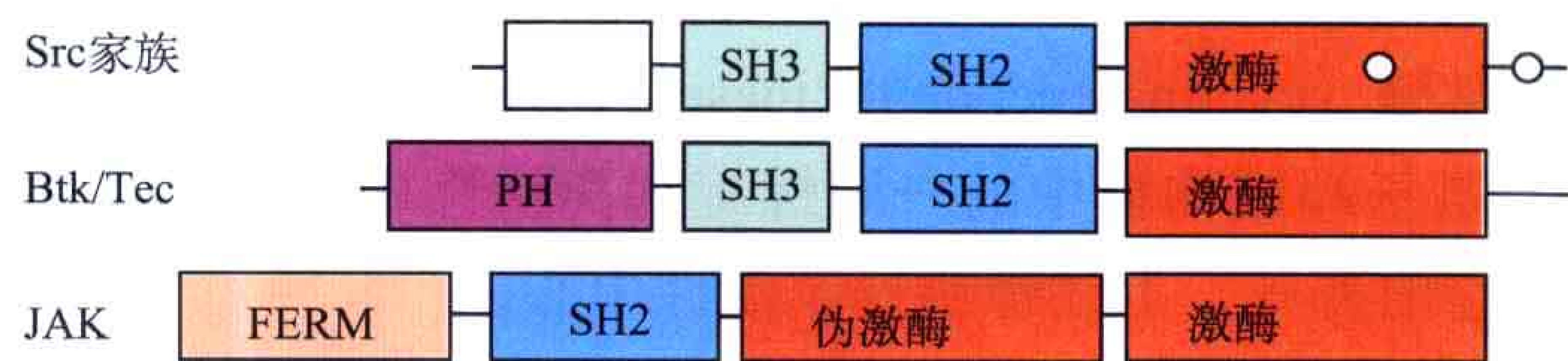


图 11.5 一些非受体酪氨酸激酶的结构域分布图。Src 在其激酶结构域和羧基端尾部含有磷酸化的位点（以小球标识）。关于 PH、SH2、SH3 和 FERM 结构域的描述见表 11.3。

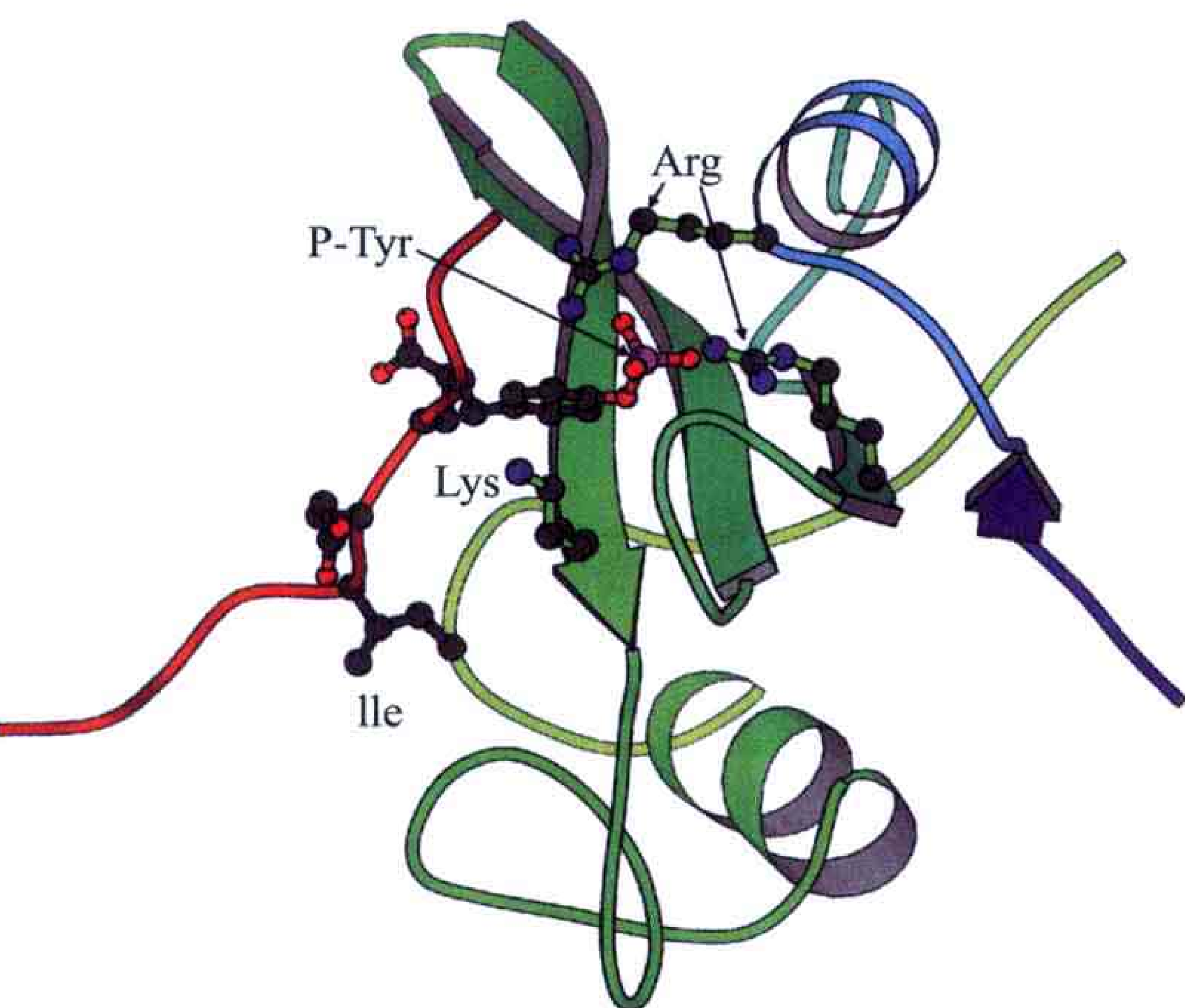


图 11.6 一个 SH2 结构域的模式图：src 蛋白的 SH2 结构域与多肽 YEEI（红色）结合的复合体。两个精氨酸（arginine）和一个赖氨酸（lysine）参与磷酸化酪氨酸（p-Tyr）上磷酸基团的结合（PDB：1SPS）。Ile 为异亮氨酸。

SH2 结构域的功能通常是特异性地结合含有磷酸化酪氨酸残基的蛋白质区段。这样，当一个蛋白质被磷酸化后，SH2 结构域就可以结合上来。

一个 SH2 结构域主要包含一个反向平行的片层结构，在片层结构的两端各有一螺旋（图 11.6）。肽段结合位点由片层与螺旋之间的部分组成。有两个口袋（pocket）可以用于底物结合：一个用于结合磷酸化的酪氨酸，另一个用于结合 Y+3 残基（见 YEEI 这个例子）。不同的 SH2 结构域含有不同的底物结合特异性，该性质主要是由口袋附近残基的性质决定的。

SH3 结构域比较小，拥有由 5 个折叠束组成的反向平行 β 桶结构（图 11.7）。与 SH2 结构域相似，SH3 功能是特异地结合其他蛋白质。被结合的蛋白质的一些区域结合在 SH3 两个卷曲形成的沟槽（groove）中。这个结构域结合的肽段的序列通常是 RXLPPLPXX 或 XXXPPLPXR。这些富含脯氨酸的肽段以多聚脯氨酸螺旋（polyproline helix）的结构存在于复合物中。此类结构的结合力相对较弱，因此在一些情况下需要多个 SH3 结构域参与结合以提高亲和力。

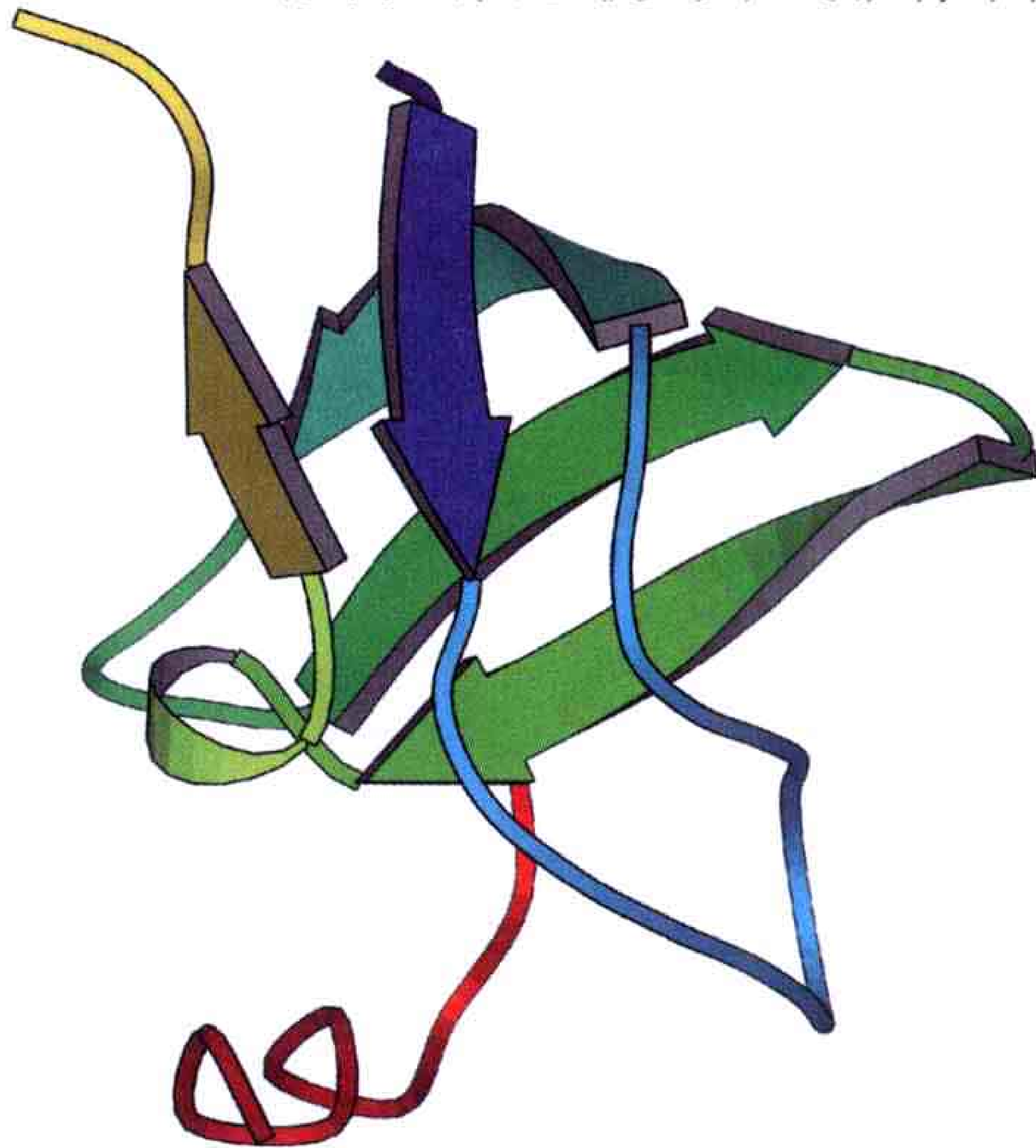


图 11.7 SH3 结构域（自于 Abl 酪氨酸激酶）与一段多肽结合的模式图。结合在 SH3 上的肽段（红色）具有 APTMPPPLPP 的序列，并形成了多聚脯氨酸螺旋的构象（PDB：1ABO）。

Src 酪氨酸激酶和一些相关的激酶含有一个羧基端尾巴，该尾部有一个可以被磷酸化的酪氨酸残基。当该残基被磷酸化时，它可以结合到 SH2 结构域上的磷酸

化酪氨酸口袋里，但是在 SH2 的第二个口袋上是不会发生结合的，因此该结合力比较弱。这个尾巴对于该蛋白质的活性调控十分重要，当酪氨酸被磷酸化后，蛋白质处于失活状态。人们发现 src 蛋白的一个致瘤性突变体存在于劳氏肉瘤病毒（Rous sarcoma virus）中，该病毒会引起鸡肿瘤的产生。这个蛋白质被命名为 v-src。这个蛋白质缺少了羧基端的尾巴，因此这个致癌基因的激酶活性不能像正常的 src 蛋白那样受到调控。这就导致了鸡肿瘤的发生。

在 SH2 结构域与 src 酪氨酸激酶之间的连接区域（linker region）含有一些脯氨酸残基，这个区域可以结合在 SH3 结构域上。失活的 src 蛋白在三维结构上拥有多个结构域相对紧凑结合的构象（图 11.8）。

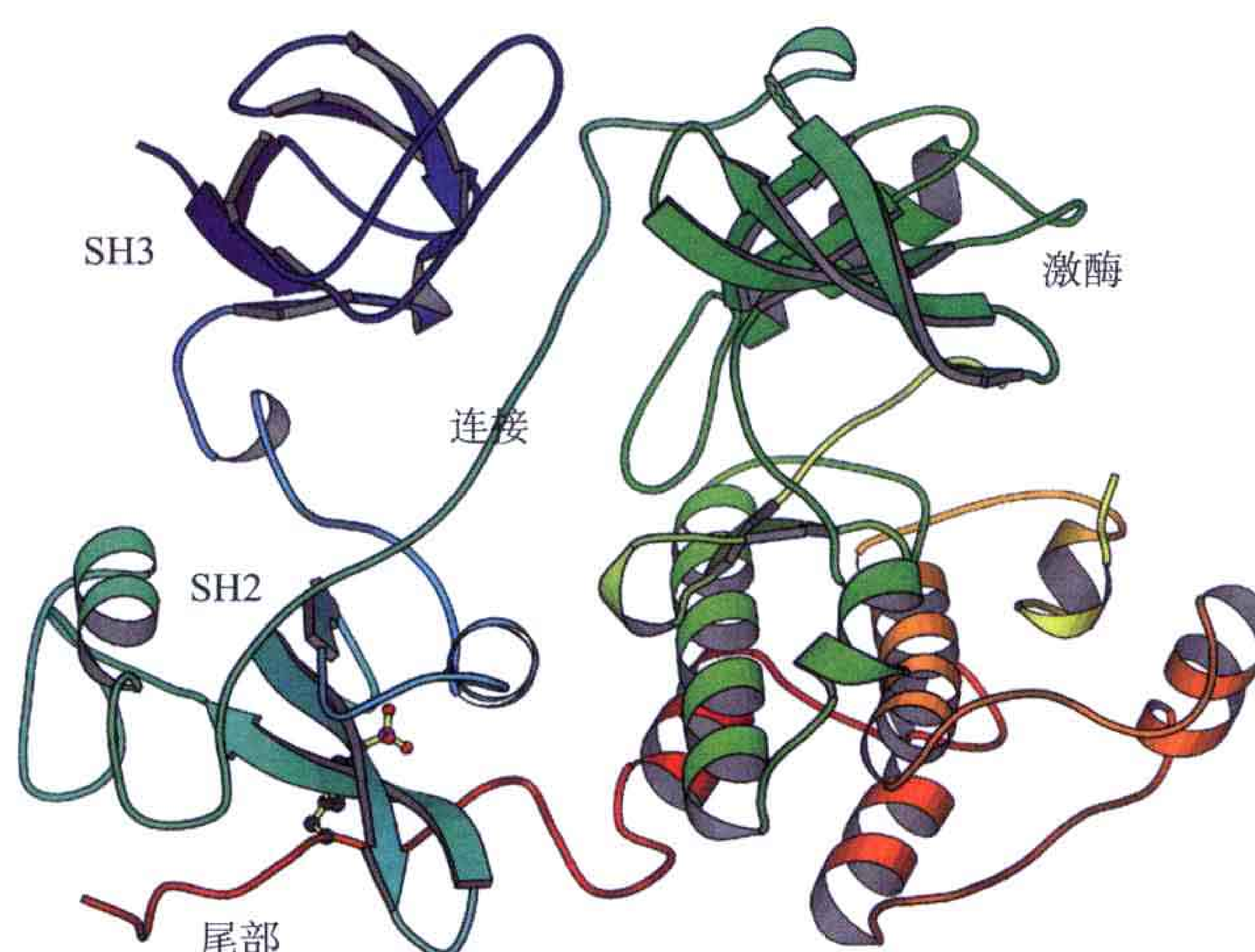


图 11.8　src 酪氨酸激酶的结构。在这个失活的状态下，SH2 结构域结合在尾部磷酸化的酪氨酸上，SH3 结构域结合在连接区。在活性状态下，SH2 和 SH3 结构域处于可以结合其他分子的状态。

该酶的激活机制还不是十分清楚。与其他的激酶进行比较后发现，src 蛋白的激酶结构域在失活的状态下， 活性中心（第一个结构域的螺旋）的部分残基处于一个抑制其活性的方向。尾部与 SH2 结构域的结合以及连接区与 SH3 结构域的结合使得这个失活构象得以稳定；而 SH2 和 SH3 结构域结合在与活性中心相对的两端。当尾部的酪氨酸去磷酸化，或其他肽段与 SH2 及 SH3 结构域发生高亲和力结合时，src 蛋白结构域之间的结合关系就会被打破，并允许该分子中的激酶结构域形成一个活性的构象。当该激酶 T-loop 的酪氨酸被磷酸化后，激酶被完全激活。

11.3　受体酪氨酸激酶通路

一大类信号通路利用本身具有酪氨酸激酶活性的受体参与信号转导。这类受体分子被称为受体酪氨酸激酶 (receptor tyrosine kinase，RTK)，其中包含了重要的胰岛素受

体和大量的生长因子受体。与细胞因子受体相比，该类受体蛋白本身含有酪氨酸激酶活性，而细胞因子受体中，信号会激活其他独立存在的激酶。

受体酪氨酸激酶包含一个胞外受体区、一个跨膜的螺旋区及胞质部分的酪氨酸激酶结构域。受体的胞外部分或者是富含半胱氨酸富含的域与亮氨酸重复结构域的组合，或者是含有不同拷贝数目的免疫球蛋白类结构域。一些受体分子在含有常见结构域的基础上还会含有一些特定的结构域。一些具有代表性的受体酪氨酸激酶的结构域示意图见图 11.9。

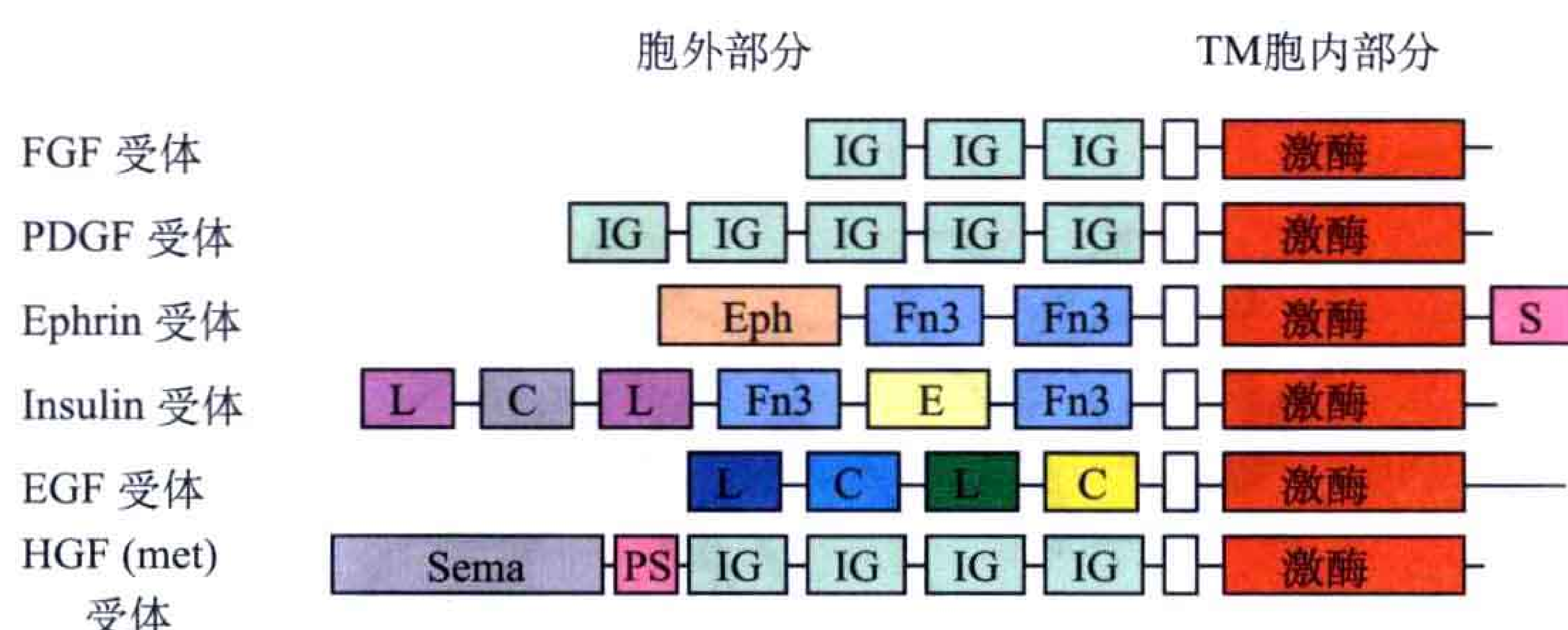


图 11.9 一些受体酪氨酸激酶的结构域分布图。它们都含有相似的胞内激酶结构域和跨膜区域（transmembrane region，TM）（由一小块白色的四方形标示）。其他的模块是免疫球蛋白结构域（IG）、Ⅲ型纤连蛋白结构域 (Fn3)、亮氨酸重复结构域 (L)、弗林蛋白酶或半胱氨酸富含结构域 (C)、SAM 结构域 (S) 及蝶素配体结合结构域 (Eph)。胰岛素受体含有一些其他受体分子中没有的结构域，肝细胞生长因子受体或甲硫氨酸受体则含有一个大的七瓣螺旋桨结构的 Sema 结构域，其后有一个 PSI 结构域。

所有受体都具有一个共同的特点，即配体的结合会激活酪氨酸激酶的活性。在大多数酪氨酸激酶受体中，配体结合到受体的胞外部分会引起受体两条链的二聚化。在这一点上，它们与细胞因子受体相似；不同点是，细胞因子受体需要另外一个单独的蛋白质来发挥酪氨酸激酶的活性。而在胰岛素受体这个例子中，没有配体的存在时，受体分子依然可以发生二聚化。配体可以直接结合到已经形成二聚化复合物的受体分子上。

当受体被磷酸化后，磷酸化的肽段就成为信号通路中其他蛋白质的靶标。这些蛋白质会利用如 SH2 结构域结合到受体酪氨酸激酶上。

11.3.1 EGF 受体通路

表皮生长因子（epidermal growth factor, EGF）是利用受体酪氨酸激酶发挥作用的一个例子。这个通路一直受到广泛关注，因为已有研究表明该受体活性的上调与癌症的发生密切相关。该受体胞内结构域的磷酸化可以激活胞内多个信号通路。这里我们仅描述 MAP 激酶通路的激活，而此通路外的 JAK/STAT 通路及蛋白激酶 B 通路也都可以被 EGF 受体激活。

表皮生长因子受体包含两个完全相同的肽链。每个单体分子含有 4 个胞外结构域、一个跨膜区、一个胞内酪氨酸激酶结构域（图 11.10）。胞外的 4 个结构域是由两个亮

氨酸重复结构域和两个半胱氨酸富含结构域（C domain）构成的。该受体一些结构已经得到解析（图 11.10）。结构域 I 和Ⅲ有时也称为 L1 和 L2（因为它们富含亮氨酸），它们拥有 β 螺旋折叠模式。β 螺旋折叠即以平行的 β 片层作为重复单元形成的短的长方形。第Ⅱ和第Ⅳ结构域是由富含半胱氨酸的重复单元组成的 C 结构域，该结构域构象是靠数个二硫键来维持的。

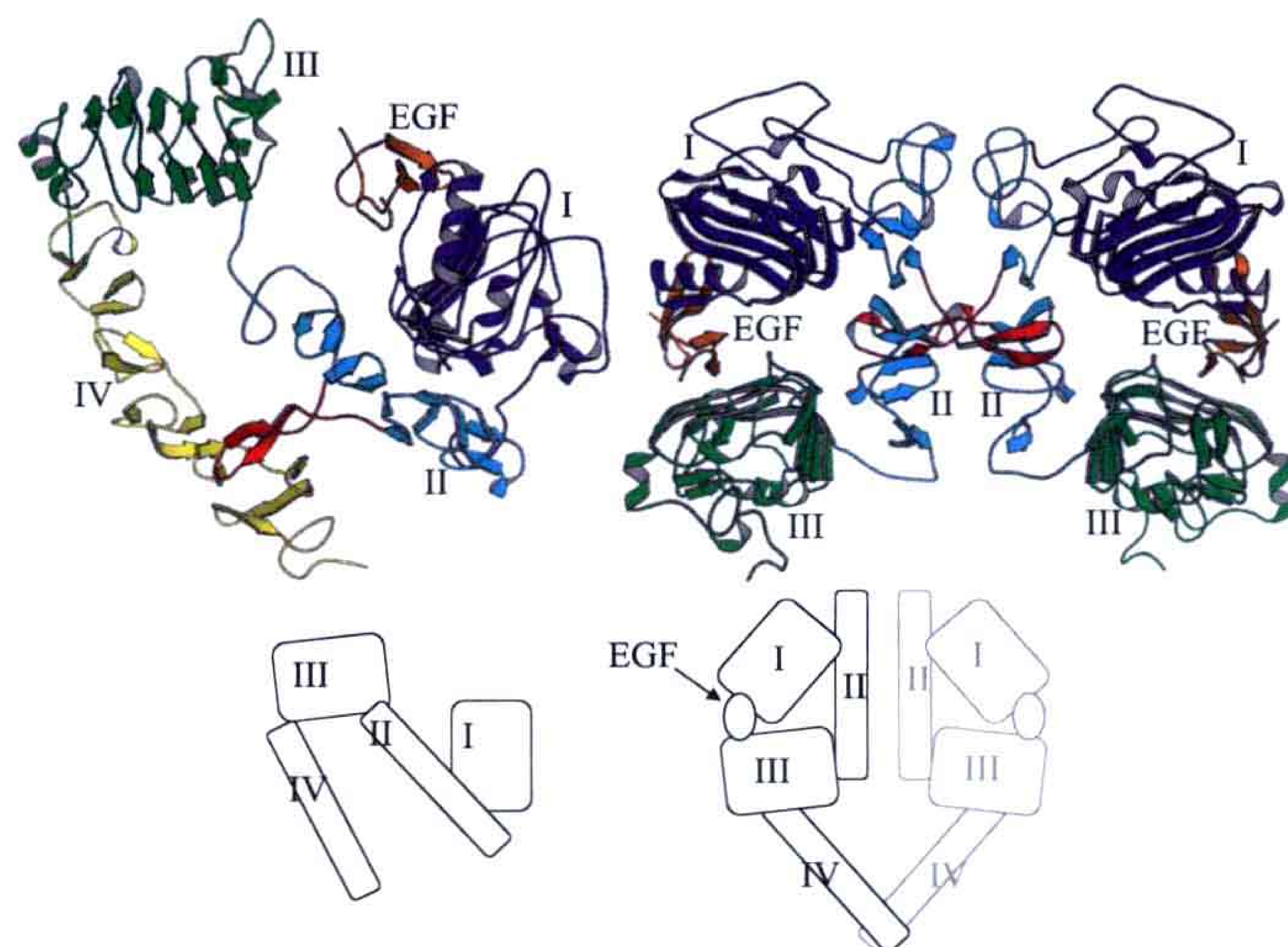


图 11.10　EGF 受体胞外部分结构。左：单体结构（PDB：1NQL）；右：二聚体结构（PDB：1IVO）。图示方向为跨膜区位于下方。在单体结构中，所有 4 个结构域都是可见的，而朝向细胞膜的结构域Ⅳ在二聚体结构中无序摆动导致在晶体结构中看不到。结构域Ⅱ中对受体二聚化起到重要作用的卷曲部分用红色标识。我们可以注意到 EGF 作为配体并没有参与受体的二聚化。在右下方的图中显示了各个结构域的排布情况，并预测了结构域Ⅳ在其中的位置。

在单体结构图中我们可以发现，4 个结构域构成一个弯曲的构象。一个生长因子分子结合在结构域 I 上。在二聚体结构中，生长因子结合在结构域 I 和结构域Ⅲ上，结构域在空间上发生了重排，与单体结构中的相对位置相比，结构域 I 与结构域Ⅱ发生了 130° 的旋转。结构域Ⅱ在受体的单体中与结构域Ⅳ发生相互作用；而在受体二聚体中，它在一定程度上通过该受体家族所特有的一段较为伸展的卷曲部分，与二聚体分子中的另外一个分子的结构域Ⅱ发生相互作用。而在其他类的受体二聚体中，配体并不是直接作用于二聚体的相互作用面上。配体（生长因子）的结合实际上稳定了由于二聚化而形成的较为伸展的构象。在受体的单体晶体结构中，因为实验中所使用的低 pH 环境，受体的构象变化无法被引发。

我们知道 EGF 结合到受体上会导致受体胞内区自磷酸化。而在二聚体结构中，连接跨膜区域的胞外结构域Ⅳ在晶体结构中却处于无序的状态，这样受体的胞外结构域究竟是如何调节胞内的酪氨酸激酶结构域也便不得而知。EGF 受体的磷酸化发生于的羧基末端（图 11.10）而并非激酶结构域本身。激酶结构域在没有配体结合时看似仍处于活化状态（图 11.11）。在其他受体酪氨酸激酶中，自磷酸化常常发生在可以调节酪氨酸激酶活性的活性卷曲部分。

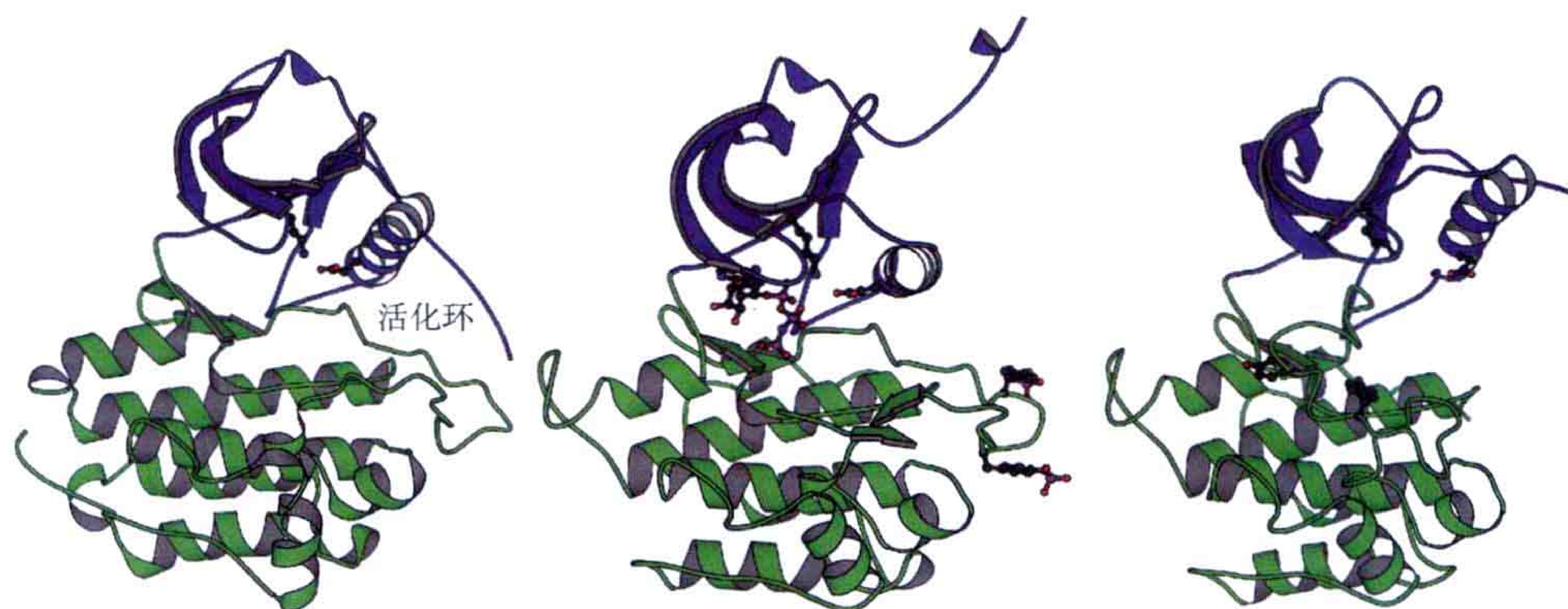


图 11.11 EGF 受体的酪氨酸激酶结构域（PDB:1M14）及其与胰岛素受体相应结构域的激活与非活性状态的比较（PDB:1IR3 和 1IRK）。

11.3.2 G 蛋白 Ras 和 MAP 激酶通路的激活

包括表皮生长因子受体通路在内的许多受体酪氨酸激酶信号通路，利用 G 蛋白 Ras 来激活一类被称为 Map 激酶的激酶级联反应（图 11.12）。这个通路反应的第一步是 Grb2 蛋白 (growth factor receptor binding protein-2, 生长因子受体结合蛋白 -2) 与自磷酸化的 EGF 受体的羧基端相结合。Grb2 蛋白中间为一个 SH2 结构域，SH2 结构域两端各有一个 SH3 结构域。SH2 结构域可以与受体上磷酸化的酪氨酸残基结合。Grb2 蛋白本身不具备酶活性，在这里它只是作为一个接头蛋白发挥作用。与其他 SH3 结构域类似，Grb2 蛋白的 SH3 结构域结合在一段脯氨酸富集的序列上，在这里，它结合在 Sos（son of sevenless）蛋白上。Sos 蛋白通过充当鸟苷酸置换因子（G nucleotide exchange factor, GEF），从而激活膜结合 G 蛋白 Ras（图 11.12）。

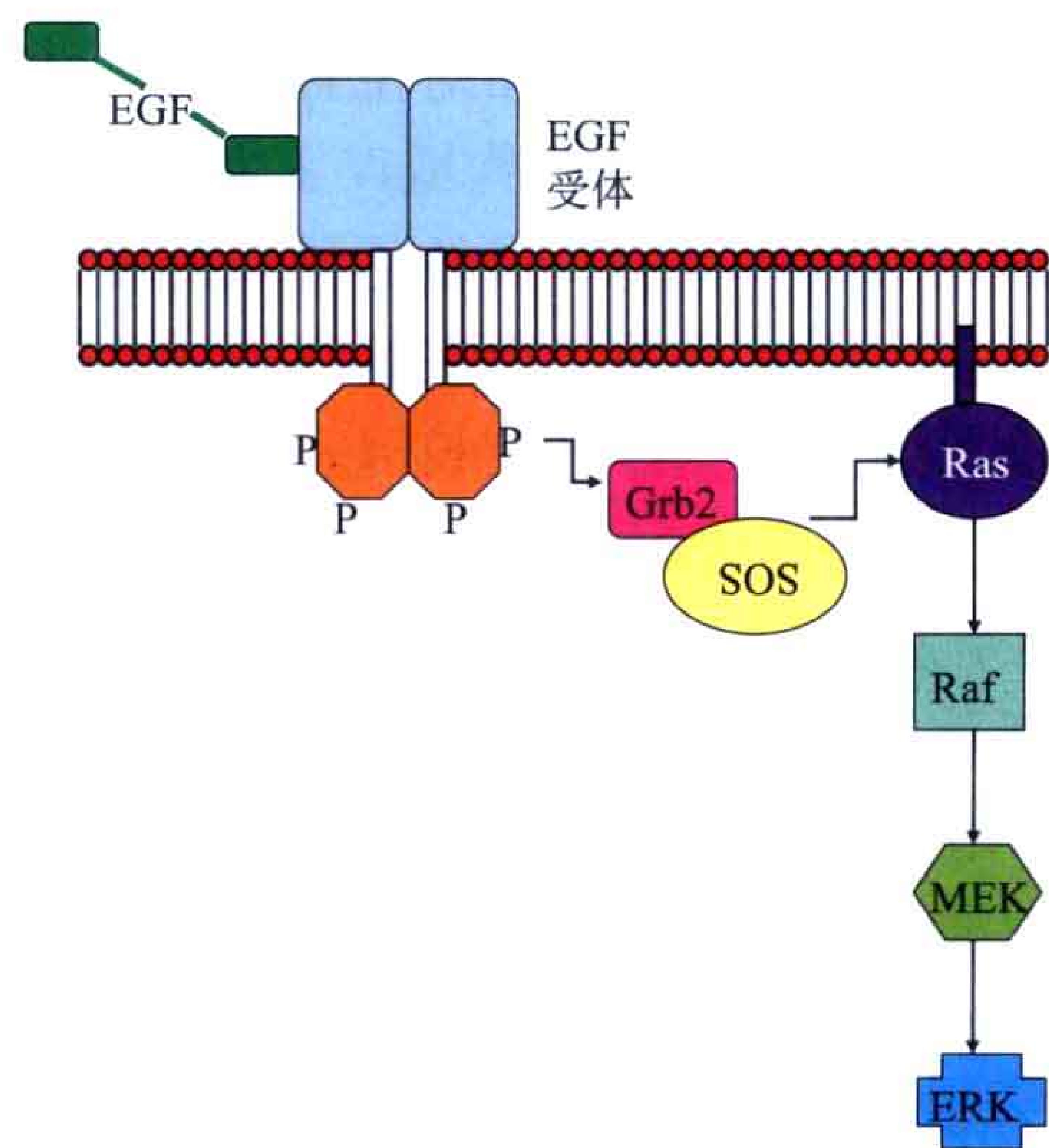


图 11.12 通过 EGF 受体激活 Ras 通路的示意略图。

Ras 蛋白的效应分子是 Raf 蛋白，它是含有 Ras 蛋白结合结构域的一个激酶。被激活的 Raf 蛋白磷酸化另一个激酶的丝氨酸残基，该激酶为 MapKK（也称为 Mek）。这个激酶反过来磷酸化激酶 MapK（又称为 ERK）。该激酶穿过核膜进入细胞核，并磷酸化激活转录因子。通过这一过程，EGF 与受体的结合就导致了某些特定基因的表达变化。

11.3.2.1　通过 GDP 释放引起的 Ras 蛋白的激活

所有的 G 蛋白都含有一个常见的 G 结构域（见第五章）。Ras 蛋白是含有该结构域的最简单的一种，Ras 蛋白也常被归为最典型的 G 蛋白（图 11.13）。

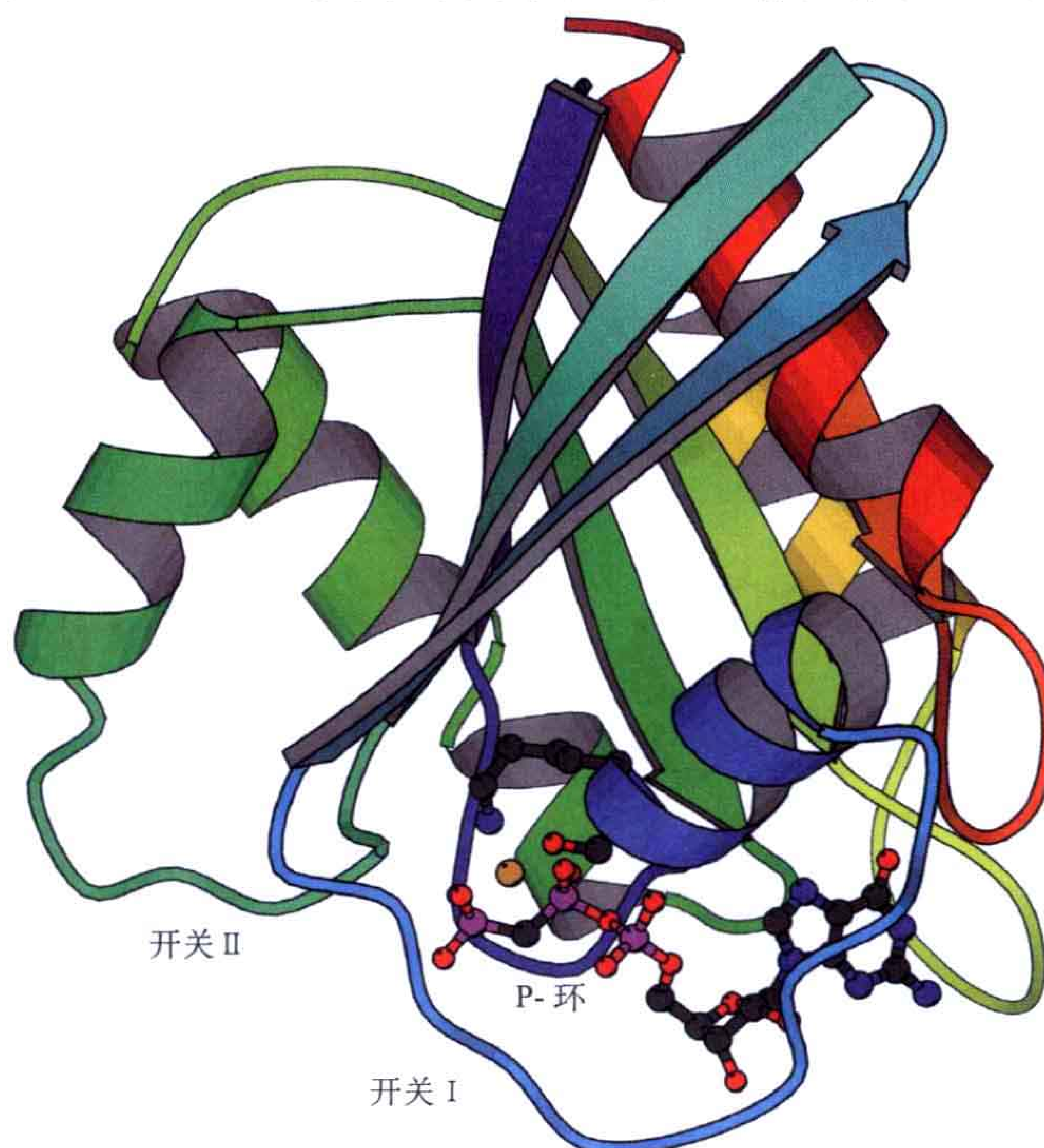


图 11.13　GTP 结合状态下的单体 G 蛋白 Ras 结构（PDB：121P）。P 环结构（P-loop）中 GXXXXGKS 模体（motif）上的第 16 位赖氨酸和第 17 位丝氨酸进行了加亮处理。赖氨酸结合到了 GTP 的第 3 个磷酸基团上，而丝氨酸的侧链与镁离子（橙色的小球）发生相互作用。

Ras 蛋白通过翻译后修饰在羧基端的半胱氨酸残基上添加了一分子的脂肪酸，该分子可以介导 Ras 蛋白结合到质膜上。Ras 蛋白的主要激活蛋白，Sos 蛋白，是一个拥有约 1400 个残基的较大的蛋白质。Sos 蛋白中参与 Ras 的 GEF 活性的区域为 560 ~ 1050 位的残基，这段区域是由螺旋构成的，含有两个结构域，其中的一个结构域参与了 Ras 蛋白中 switch 区域的结合。这两部分的结合面积较大（图 11.14）。Sos 中的一端螺旋发夹（helical hairpin）结构插入 Ras 蛋白中 switch Ⅰ与 switch Ⅱ之间（图 11.15）。switch Ⅰ区域取代了核苷酸的结合部位，Sos 螺旋发夹中的残基侧链插入镁离子的结合位点和一部分 GDP 的结合位点，这样便启动了 GDP 分子的释放。

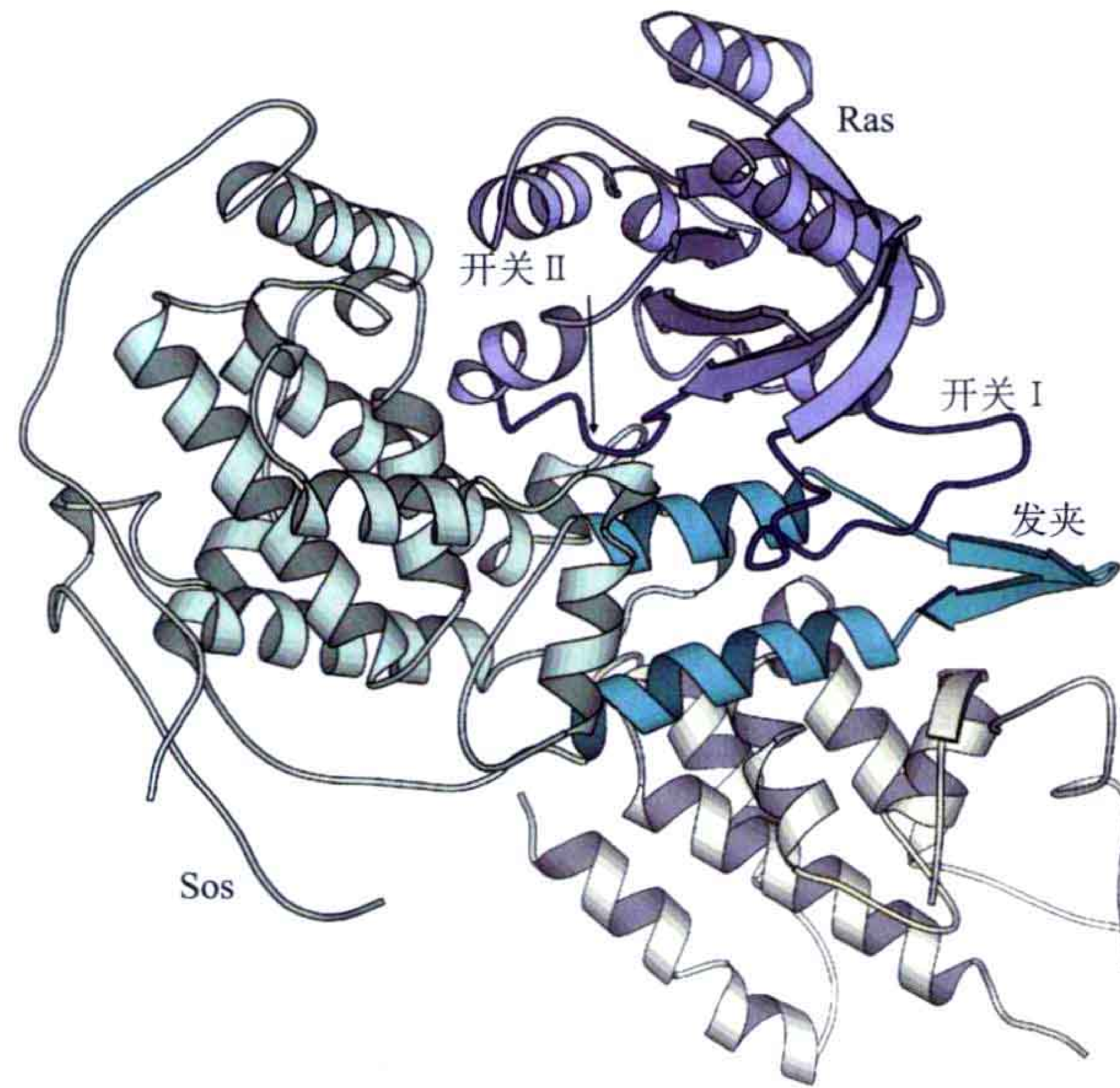


图 11.14　Sos 与 Ras 蛋白复合物结构。Sos 是一个可以结合到 G 蛋白上的 GEF 蛋白。Sos 蛋白中 Ras 结合部分包含两个结构域，即 RasGEF 结构域（白绿色）和 RasGEF 氨基端结构域（白黄色）。RasGEF 结构域中的螺旋发夹（蓝绿色）插入 Ras 蛋白（浅蓝色）的 switch Ⅰ与 switch Ⅱ之间，导致了 GDP 分子的释放（PDB:1BKD）。

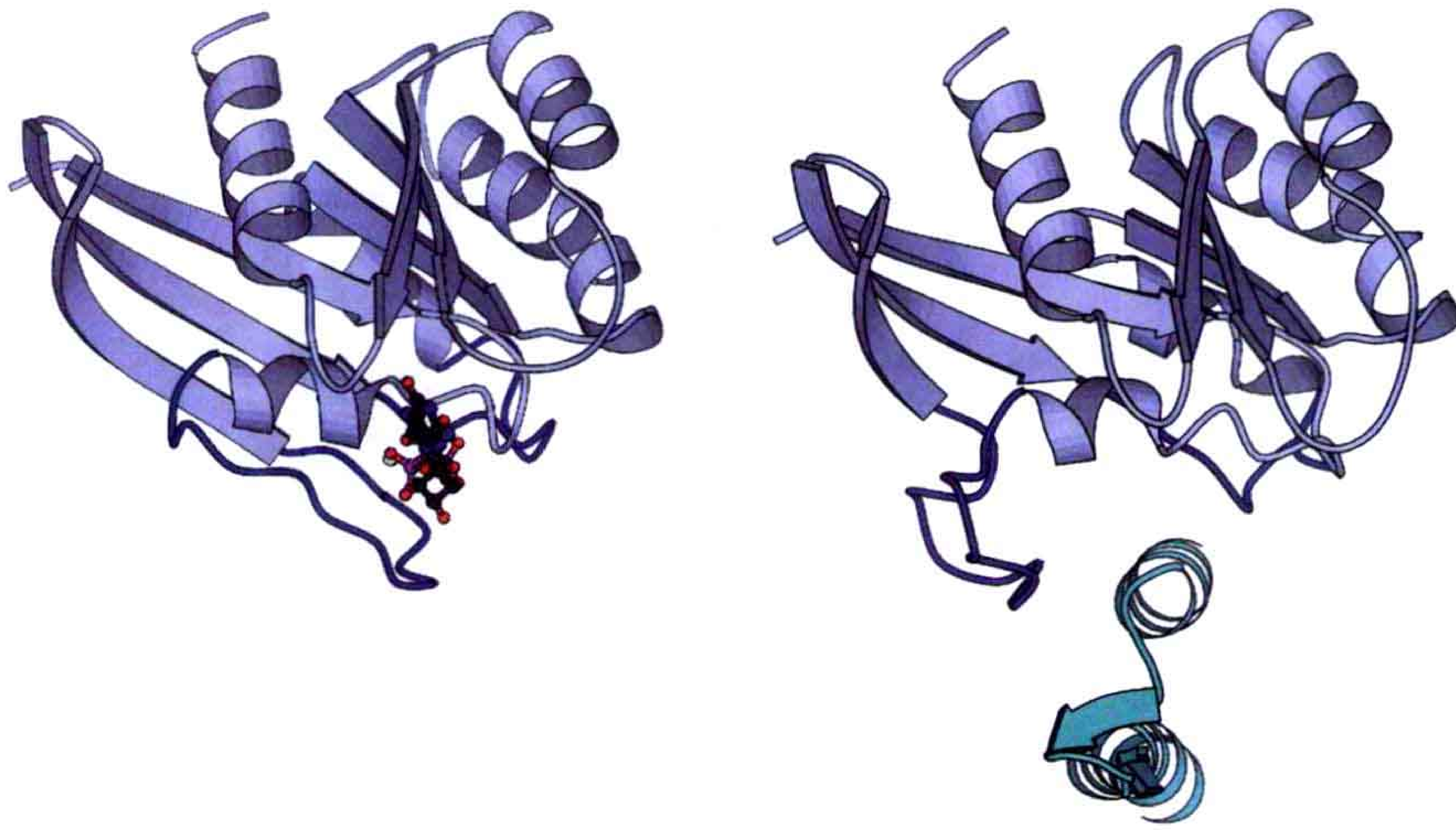


图 11.15　Ras-GDP 与 Ras-Sos 复合物结构比较。从图中我们可以看出 Sos 的螺旋发夹结构（右图下方）插入 Ras 蛋白的 switch Ⅰ（两图中的左侧部位）与 switch Ⅱ之间，并将核苷酸挤出。switch Ⅰ与 switch Ⅱ用深蓝色标识。

11.3.2.2　G 蛋白的三磷酸鸟苷酸水解酶活性和 GAP 蛋白发挥的作用

三磷酸鸟苷酸水解酶（GTPase）通常作为分子开关（molecular switch）（见 5.3.3 节）。当有 GTP 存在时，它们处在“开”的状态；当 GTP 水解为 GDP 时，它们处在“关”的状态。上述 switch Ⅰ与 switch Ⅱ的构象与这两种构象完全不同。G 蛋白的三

磷酸鸟苷酸水解酶活性通常较低。GTP 的水解机理包括氢氧根对 γ 磷酸的攻击。来自于 switch Ⅰ 的谷氨酰胺与来自于 switch Ⅱ 的精氨酸参与了该反应。这些残基通常是保守的，但是在 Ras 和相关蛋白中，精氨酸是不存在的。在 Ras 激活过程中，参与活化的活化的 GAP（GTPase activating protein, 三磷酸鸟苷酸水解酶激活蛋白）贡献了一个处在顺势(trans)的精氨酸，这个精氨酸的位置正好与其他 G 蛋白中的精氨酸位置相同，因此可以弥补 Ras 中没有精氨酸这个缺陷。而这个来自于 GAP 的精氨酸的出现将 Ras 水解 GTP 的速率提高了 1000 倍。因此通常将这个精氨酸称为“the arginine finger”。

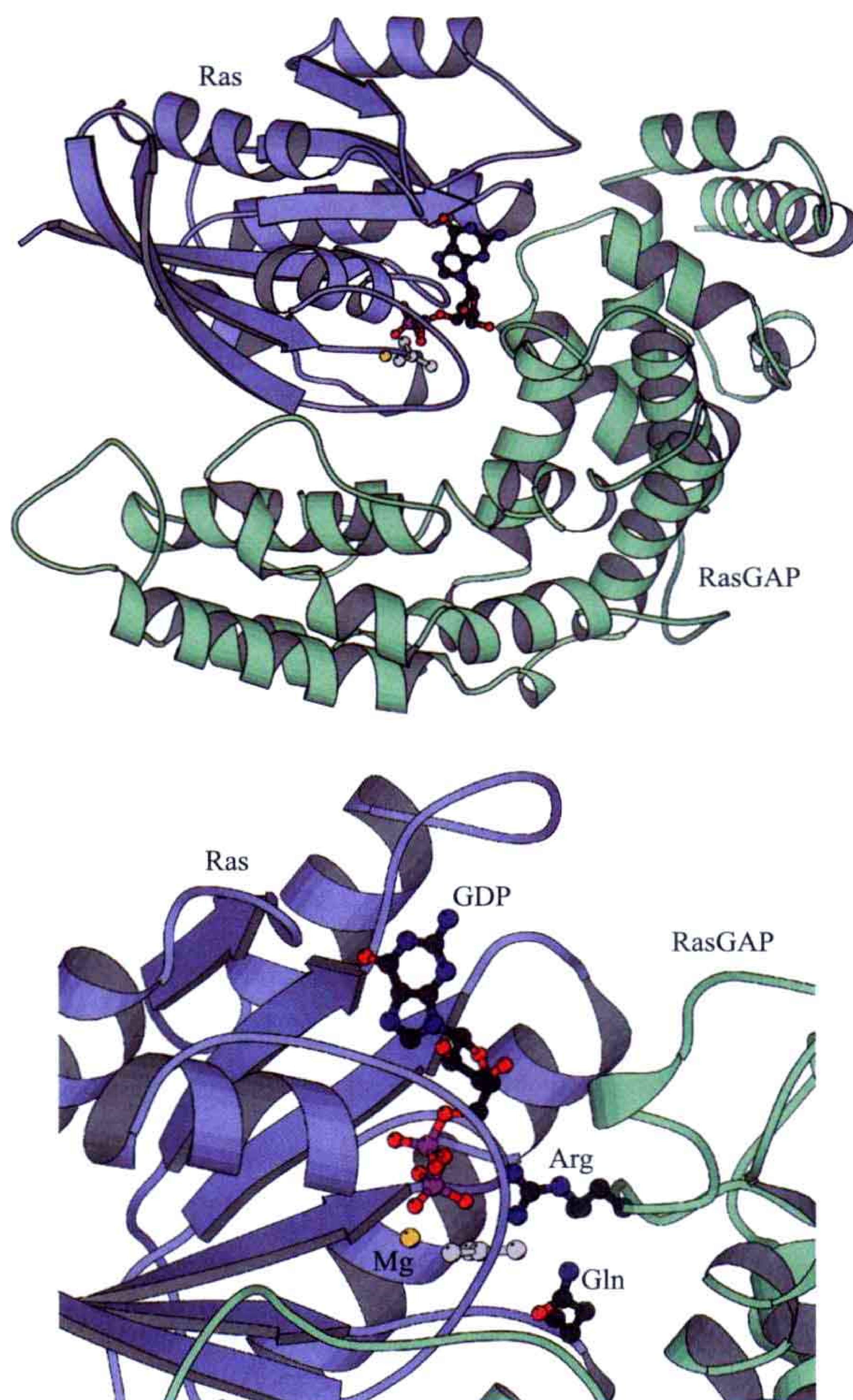


图 11.16　上：RasGAP 与 Ras 蛋白的复合物结构。RasGAP 的结合稳定了 Ras 含有部分活性中心的卷曲部分。下：Ras、RasGAP、GDP 和 AIF_3 复合物的结构细节，该图模拟了 GTP 的水解中间体状态。GAP 贡献的精氨酸（Arg）参与了 GTP 的水解（精氨酸指 arginine finger）。该图中的取向有一些细微的差别（PDB：1WQ1）。

从 RasGAP 与 Ras 的复合物结构中可以看出，RasGAP 结合在 Ras 蛋白的 switch Ⅰ 与 switch Ⅱ 区域（图 11.16）。通常情况下，GAP 蛋白对于三磷酸鸟苷酸水解酶活的贡献是稳定卷曲区域，这些卷曲区域含有参与催化后重要氨基酸。当 GAP 与 GTPase 结合后，可以使得活性中心的残基处于催化的正确位置，这些残基使得活性中心参与反应的水分子处于正确的位置并可以被活化。这些卷曲部位柔性变小所产生的负熵效应被 GAP 蛋白的结合中和掉。

11.3.2.3　Ras 蛋白和癌症

Ras 蛋白的突变与多种癌症的发生相关。这些突变通常将 Ras 永久地困在“开”的构象下而阻止 GTP 的水解。有两种突变十分常见：一种是 P-loop 上第 12 位甘氨酸到缬氨酸的突变，这个突变阻止了 RasGAP 对三磷酸鸟苷酸水解酶的激活；另一种常见的突变发生在第 61 位的谷氨酰胺，该残基可以使参与反应的水分子正确定位。

11.4 G蛋白偶联受体通路

大量的信号通路利用G蛋白和一些相似的组分在细胞中造成一些蛋白质活性的改变。这些通路利用由7个跨膜螺旋所组成的膜蛋白参与通路反应。在细胞内，受体分子与三聚体G蛋白家族中的一个成员蛋白发生相互作用。因此这类受体通常称为G蛋白偶联受体（GPCR），这个通路也被称为G蛋白偶联受体通路(GPCR pathway)。G蛋白通常与效应分子相互作用，这些效应分子通常控制cAMP及钙离子等第二信使的水平。这里我们以视觉系统为例进行描述，因为在视觉通路里有更多的结构信息。

在视觉信号的转导过程中，受体是视蛋白（opsin）。发色基团视黄醛（retinal）结合在这个蛋白质上。在视杆细胞里，视蛋白与视黄醛形成的复合物称为视紫红质（rhodopsin）。视紫红质分子对于光子十分敏感（图11.17）。当视紫红质被光激活时，视黄醛分子发生构象变化，这导致了视紫红质分子的构象变化，从而引发了三聚体G蛋白转导素（transducin）的激活。视紫红质扮演着鸟苷酸置换因子的角色，它通过诱导GDP释放和GTP的结合激活转导素。被激活的视紫红质通过激活数以百计的转导素分子而产生重要的放大步骤。被激活的转导素通过去除cGMP磷酸二酯酶的抑制性γ亚基来激活该酶。这个酶可以降解环化的GMP，从而使cGMP控制的钠离子通道（cGMP-gated Na^+ channel）关闭。通过这一步骤，转导素迅速地水解GTP分子并被视紫红质迅速再激活。因此，光子作用于视紫红质分子的视黄醛上导致了细胞膜电势的改变，而分泌的神经递质就导致了神经脉冲的发生。

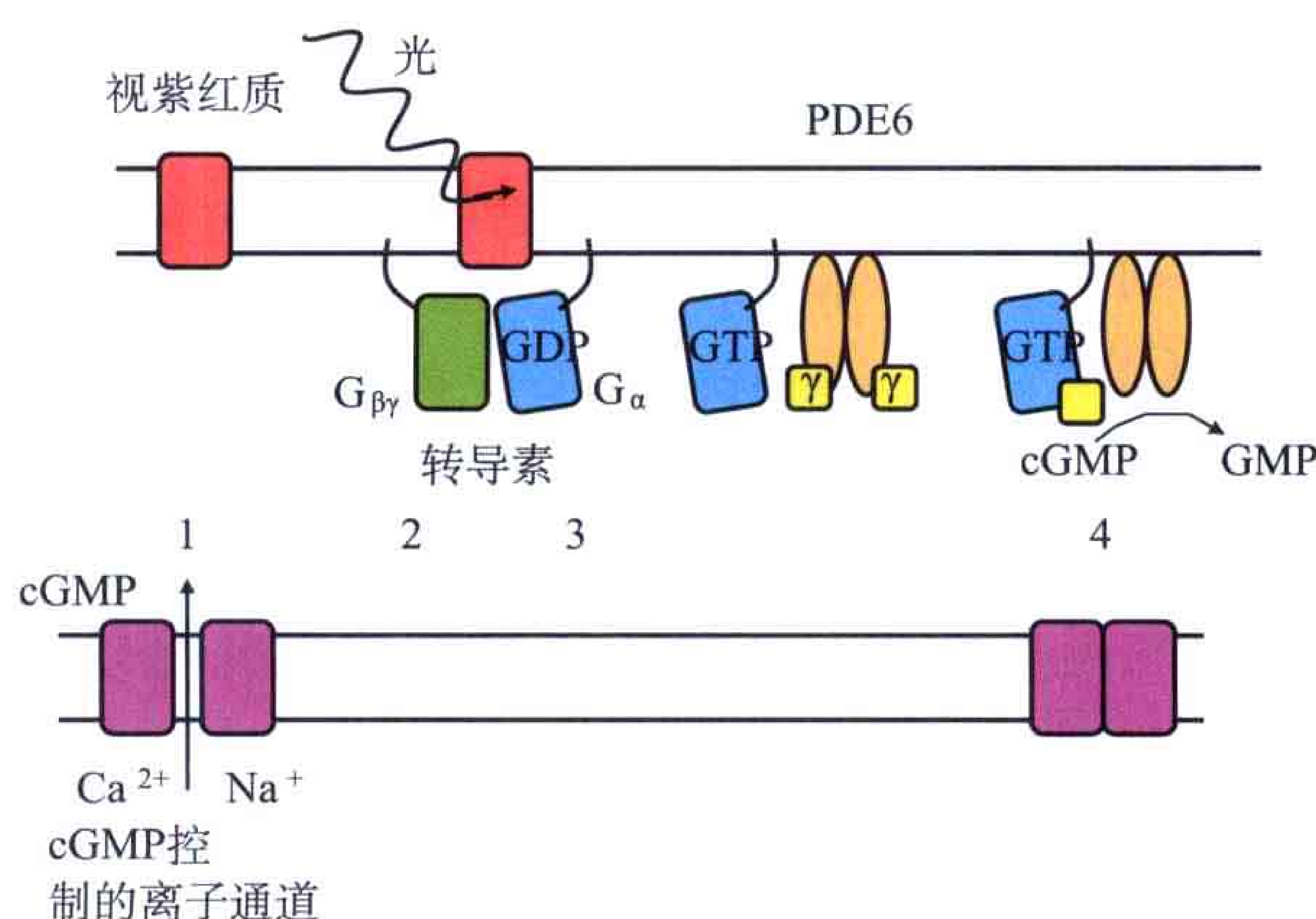


图11.17 眼视杆细胞中的信号通路略图。在没有光刺激时，cGMP使其控制的钠离子通道保持在开放状态（1）。当G蛋白偶联受体视紫红质被光子激活时，三聚体G蛋白转导素通过GDP与GTP的转换被激活（2），此时Gα亚基被释放（3）结合在环单磷酸鸟苷酸磷酸二酯酶PDE6的抑制性γ亚基上（4）。这导致了cGMP的水解并使得离子通道关闭。

11.4.1 受体

人类基因组中含有几百个不同的编码 GPCR 的基因，这使得这个蛋白质家族成为拥有大量成员的家族之一。这类受体中比较著名的有肾上腺素受体、视紫红质和各种嗅觉受体（表 11.4）。根据这些蛋白质一级序列的相似性，可以将其分成几个家族。大多数的受体属于视紫红质类 G 蛋白偶联受体家族（rhodopsin-like GPCR）。人类所拥有的其他家族有胰泌素类 G 蛋白偶联受体家族（又称为胰高血糖素受体类 G 蛋白偶联受体家族）和亲代谢型谷氨酸受体家族。值得注意的是，这些家族在蛋白氨基端位于跨膜区前的胞外部分大小有所不同。视紫红质类受体的氨基端部分比较短而谷氨酸受体的氨基端部分较大，它由 600 多个残基构成了受体的胞外表面。

表 11.4 一些 G 蛋白偶联受体通路中主要组分的举例

受体	G 蛋白	效应物
视紫红质	$G_t\alpha$（转导素）	环单磷酸鸟苷酸磷酸二酯酶
肾上腺素受体	$G_s\alpha$	腺苷酸环化酶
胰高血糖素受体	$G_s\alpha$	腺苷酸环化酶
嗅觉受体	$G_{olf}\alpha$	腺苷酸环化酶

11.4.1.1 视紫红质的结构

G 蛋白偶联受体将细胞膜外的受体结合信号转换成膜内蛋白的构象变化。视紫质蛋白可以被看做是这一类受体的代表，而与其他受体分子的唯一不同之处是视紫质无需胞外可溶性的配体结合来启动信号通路。由于完整的膜蛋白通常难以结晶，所以通常只有单独的胞外结构域得以解析。G 蛋白偶联受体在结晶上尤其难以操作，但是视紫红质（图 11.18）和结构类似的 β2 肾上腺素受体的结构最近也得以解析。

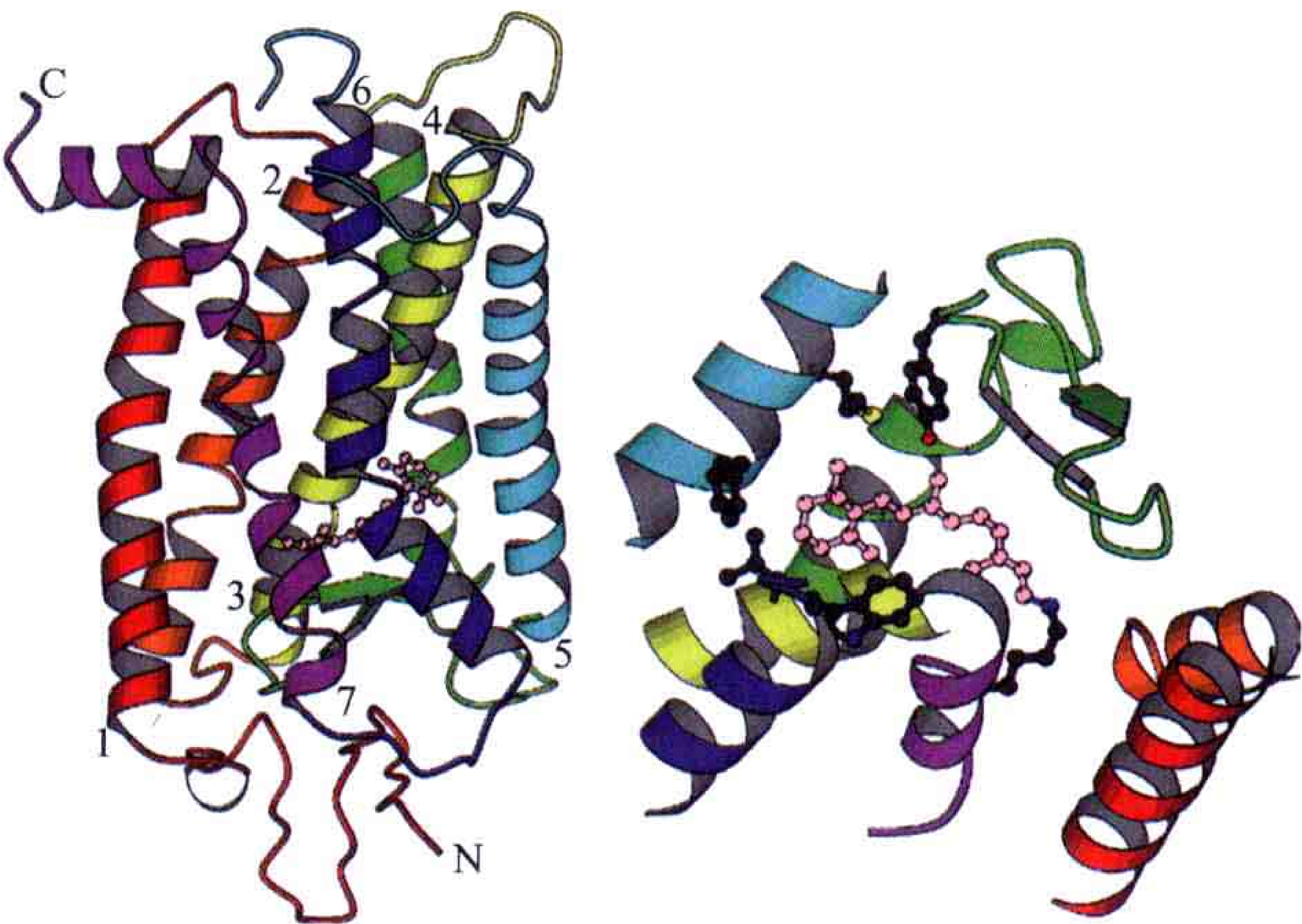


图 11.18 左：7 次跨膜螺旋的受体分子——视紫红质的结构图（与膜平面方向视图）。胞外部分位于图中的底部。胞内部分在羧基端含有一段两性螺旋，这段螺旋的一部分在晶体结构中不可见。可以看到在中间用球棍模型标识的视黄醛分子。右：视黄醛分子（粉色）的环状结构结合在一个疏水的口袋里，尾部共价连接在赖氨酸残基上（PDB：1F88）。

7 次跨膜螺旋垂直于细胞膜所在的平面并形成一束，视黄醛分子坐落在螺旋形成的通道中央。螺旋 1 与 2、3 与 4、5 与 6 之间的卷曲及羧基末端的卷曲部分暴露于胞质部分。

视黄醛分子是由一个环状结构加一个尾巴形成的碳水化合物（图 11.19）。视黄醛的尾部共价地连接在蛋白质的赖氨酸残基上。可见光能将视黄醛分子 11 位的顺式结构（*cis*）转换为反式结构（*trans*）。这影响了跨膜区螺旋的排列以及螺旋间处于胞内部分卷曲的空间排布，进而影响了与视黄醛结合的 G 蛋白转导素，但是这些变化的细节部分我们还不得而知。光敏性视黄醛分子引起的构象变化效应与其他配体和受体相结合引起的效应相类似——可以将光子看做能够共价“结合”在视黄醛分子上的“配体”。

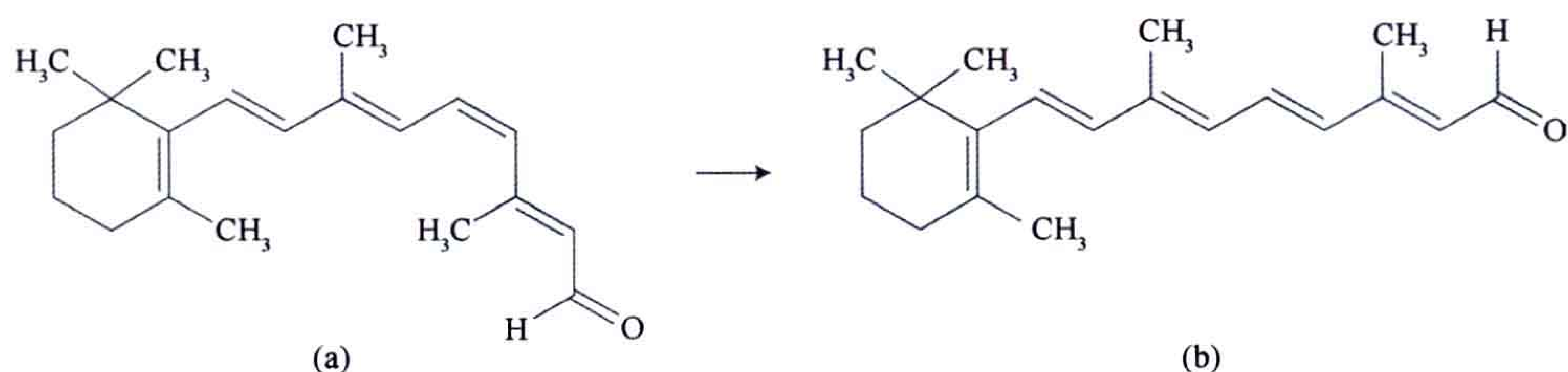


图 11.19　视黄醛分子和光诱导的其构象变化。(a) 11 位顺式视黄醛；(b) 反式视黄醛。在视紫红质中，该分子共价结合到赖氨酸残基上。

细菌中，有很多光敏感蛋白在三维结构上与视紫红质很相似。其中一个蛋白是细菌视紫红质，它是细菌中一个光驱动的质子泵。对该蛋白质的研究有助于人们了解由光吸收引起的蛋白质构象变化的起始阶段。细菌视紫红质也会结合一个类似视紫红质中的视黄醛分子。通过将细菌视紫红质蛋白晶体维持在低温下并用光对其进行照射，光介导的转换中间步骤用 X 射线晶体学进行研究变得可能，这样就可以显示出这些变化在原子水平上的信息。但是我们不知道细菌视紫红质的这些变化与视紫红质发生的光诱导变化到底在多大程度上是一致的。

11.4.2　三聚体 G 蛋白：分子开关

三聚体 G 蛋白在 G 蛋白偶联受体通路中担当着信号放大者的角色：当其被激活时，如结合到活化的受体分子上，它们将许多处于失活状态的效应物激活。三聚体 G 蛋白包含三条链，即 α、β 和 γ 链（图 11.20）。三聚体 G 蛋白利用与 α 亚基的氨基端、γ 亚基的羧基端共价结合的脂质分子锚定在细胞膜上。由于被激活的受体分子和受三聚体 G 蛋白调控的效应物分子都是结合在细胞膜上的，所以 G 蛋白结合到膜上可以提高它的效率。

结合 GDP 的非活性三聚体 G 蛋白结合于受体分子上，该受体通常与视紫红质一样属于 7 次螺旋组成的跨膜蛋白。受体分子有一个静息状态，当配体与其结合时（如肾上腺素与肾上腺素受体结合、光子刺激视紫红质等），它会发生构象变化。被活化的

受体分子之后便会发挥其 GEF 蛋白的活性，包括 GDP 分子从非活性 G 蛋白上释放和 GTP 与活性 G 蛋白结合。

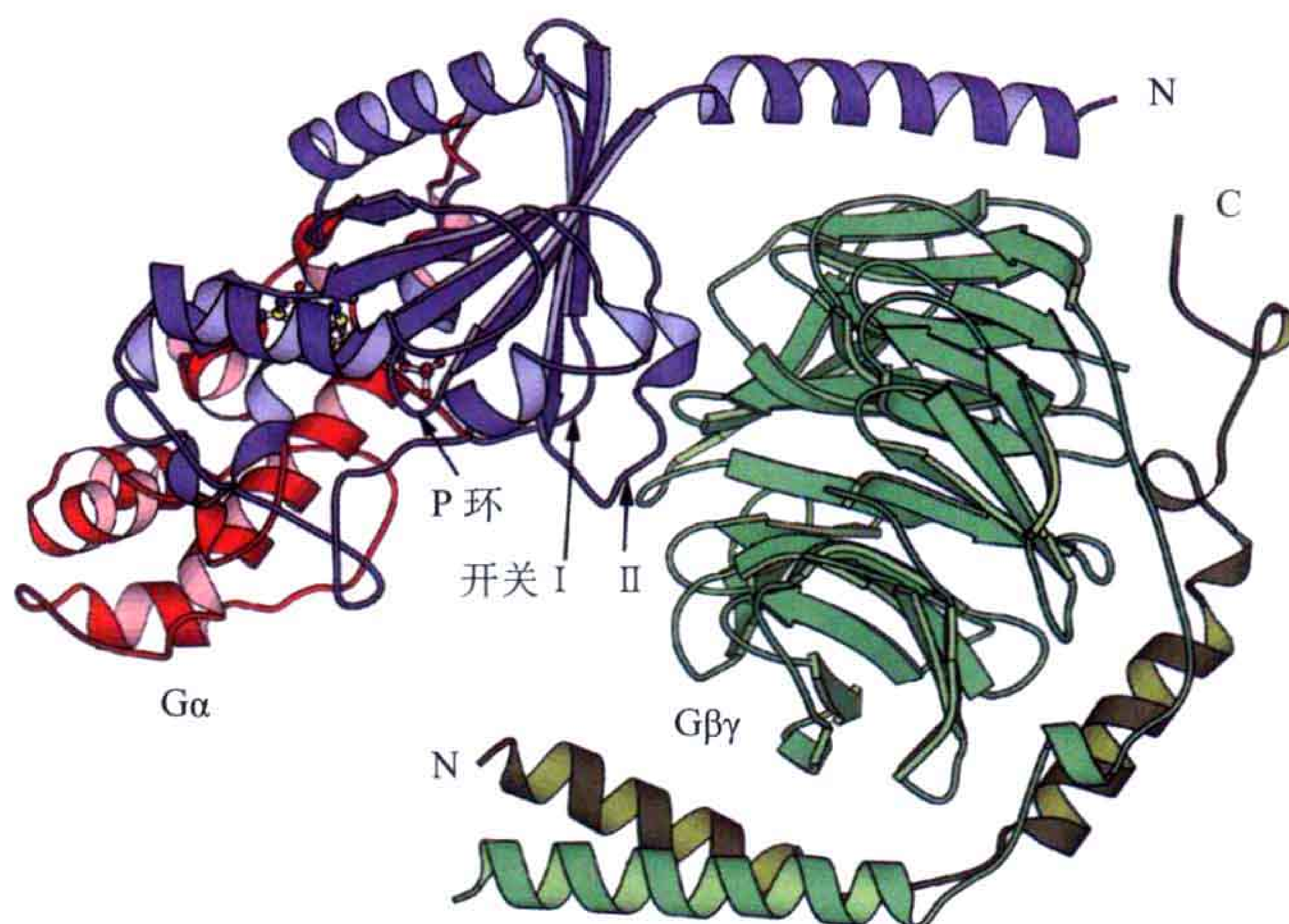


图 11.20　$G_{\alpha\beta\gamma}$ 复合物结构（转导素分子，PDB：1GOT）。α 链拥有 G 结构域（蓝色）和一个插入其中的螺旋结构域（红色）。$G_{\beta\gamma}$（G_β 显示为绿色，G_γ 显示为橄榄色）结合到 G 结构域的 switch 区域。

当 GTP 结合时，激活的 G 蛋白分为两个部分，即 G_α 和 $G_{\beta\gamma}$。这两个分开的部分对于受体分子的结合力比较小，使得其可以从受体分子上脱离并发生移动，这样可以激活和影响效应蛋白。只有当 GTP 水解为 GDP 时 G 蛋白才会失活，因此信号是否传递依赖于 G 蛋白三磷酸鸟苷酸水解酶的活性。当 GTP 水解成 GDP 时，G_α 和 $G_{\beta\gamma}$ 重新形成三聚体结合到受体分子上，使得下一轮的激活可以发生。

三聚体 G 蛋白的 G_α 亚基含有两个结构域：一个是三磷酸鸟苷酸水解酶结构域，另一个为螺旋结构域（图 11.21）。三磷酸鸟苷酸水解酶结构域是一个与 Ras 相似的经典的 G 结构域。螺旋结构域较长，插入 G2 结构域或 Ras 折叠的 switch 卷曲处（图 11.20）。这个结构域像一个盖子盖住了与蛋白质结合的核苷酸。

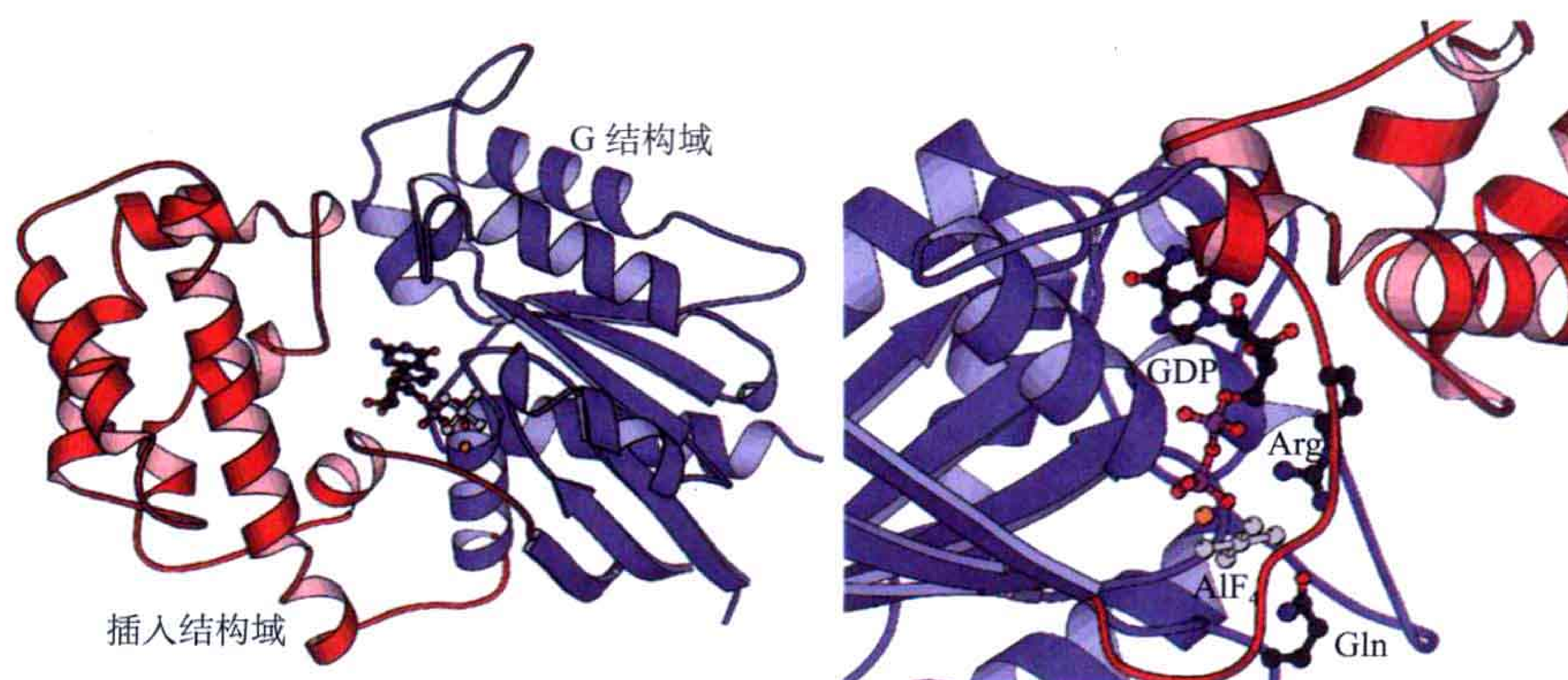


图 11.21　左：转导素 α 亚基的 G 结构域和插入的螺旋结构域（PDB：1TAD）；右：螺旋结构域作为一个内部的 GAP 蛋白，贡献一个精氨酸（the Arg-finger）到活性中心。含有 GDP 与四氟化铝的结构模拟了 GTP 水解成中间产物的中间态结构。参与固定活性水分子的谷氨酰胺残基也用球棍模型显示出来。

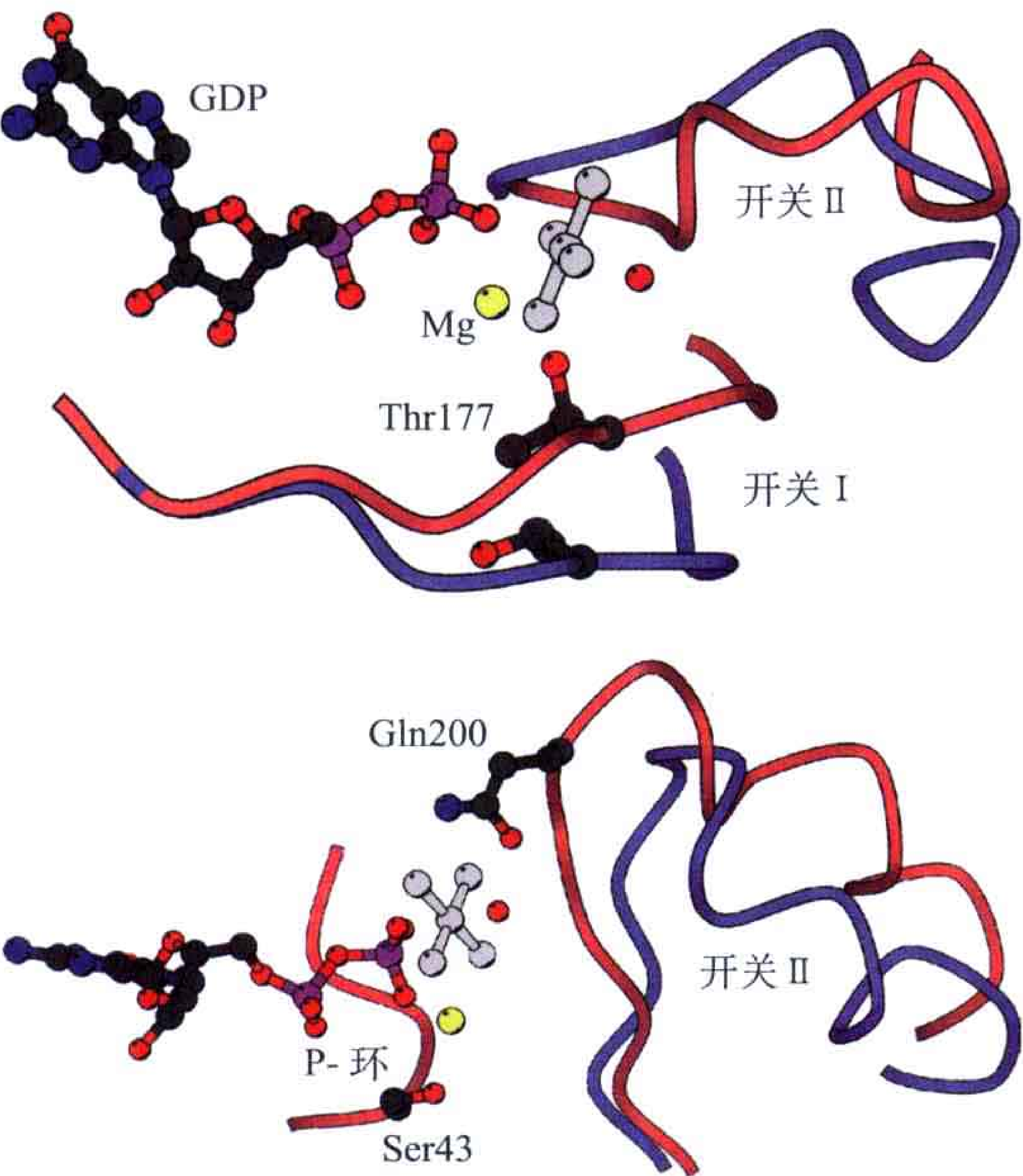


图 11.22 镁离子和 γ 磷酸附近的构象变化。上：在 GTP 结合构象中，177 位的苏氨酸结合到镁离子和 γ 磷酸上（红色）。在其他的 GTPase 蛋白中，GTP 的结合结构通常是由 GTP 类似物结构或中间态结构来得到的，在这个例子中是利用 GDP 和 ALF_4^- 模拟中间态。蓝色的部分为 switch Ⅰ和 switch Ⅱ在 GDP 结合状态下的构象。下：在结合 GTP 类似物的过程中，由于与“第三个磷酸基团”的氧原子形成了新的结合，switch Ⅱ发生了构象变化。这导致了 switch Ⅱ螺旋的移动。

G_β 是一个 7 瓣状的螺旋桨结构蛋白，每一个瓣状结构是由 4 个反向平行的 β 折叠束所形成的 β 折叠片层结构。它的氨基端含有一段螺旋结构，其后是由 36 ~ 40 个残基组成的 7 个重复的模体，这个模体常常以色氨酸或天冬氨酸结尾（WD）。第一个 β 折叠束处在第 7 个即最后一个螺旋片瓣的最外侧，其中保守的残基对于该结构的稳定性发挥着重要的作用。G_γ 与 G_β 分开时，形成一小段无序的结构，但当它与 G_β 结合时以两端螺旋结构构成整个蛋白质的一部分。

G 蛋白的 G_α 亚基通过包括 switch 在三聚体 G 蛋白的螺旋插入结构域（图 11.21）贡献一个参与三磷酸鸟苷酸水解酶活性的精氨酸（Arg-finger）。这个螺旋结构域因此可以被看做是一个内部的 GAP 结构域。即使通过基因工程手段将这个结构域从蛋白质中分开而作为一个单独的蛋白质分子放入细胞中，它依然可以发挥其应有的活性。

G 结构域含有三个部分，可以依据结合 GTP 或者 GDP 而发生构象变化。其中的两个部分与 Ras 蛋白 switch Ⅰ和 switch Ⅱ相同。由 γ 磷酸的存在与否引起的构象变化是颇为显著的。在 GDP 结合构象中，来自于 P-loop 上的丝氨酸和 β 磷酸上的氧原子与镁离子结合。在 GTP 结合构象中，switch Ⅰ上的磷酸和 γ 磷酸上的氧原子也会参与镁离子的结合。γ 磷酸的存在会使 switch Ⅰ向镁离子和磷酸的方向移动（图 11.22）。

switch Ⅱ的构象变化是由它和 γ 磷酸的直接相互作用造成的。switch Ⅱ上的一个主链氮原子与磷酸通过氢键相互作用，改变了由该卷曲所连接的 β 折叠束和螺旋的方向。

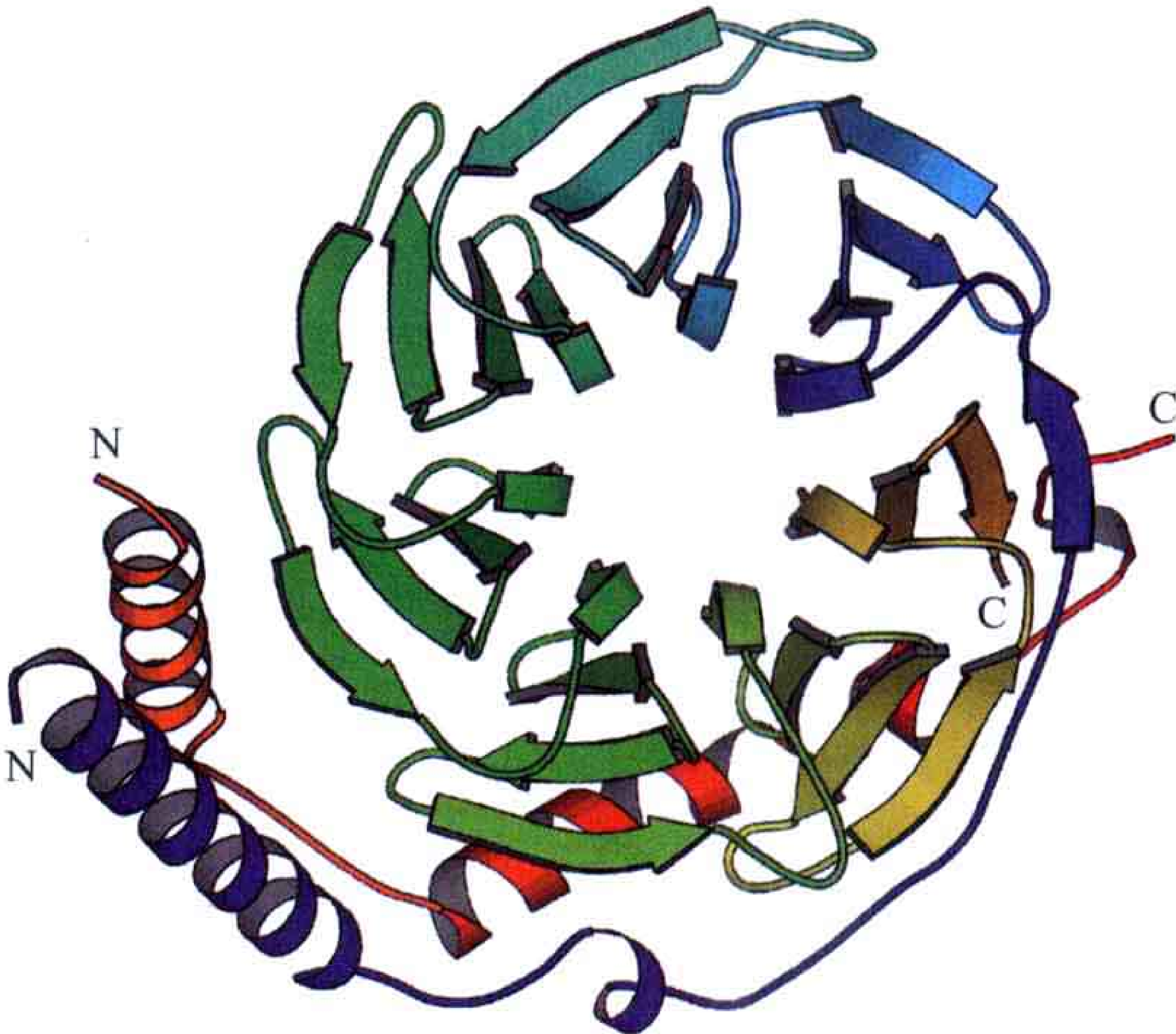


图 11.23 视导素中 $G_{\beta\gamma}$ 复合体。G_γ 显示为橙色和红色。

内的一部分与 $G_{\beta\gamma}$ 结合。结合发生后锁定了 GDP 构象中的 switch 区域（图 11.23）。G_{α} 亚基氨基端原来无序折叠的部分也变得有序起来，形成一段与 $G_{\beta\gamma}$ 相结合的螺旋。$G_{\beta\gamma}$ 的作用是稳定 G_{α} 亚基的 GDP 结合非活性状态。当 GPCR 受体引发 GDP 的释放后，GTP 的结合会导致 G_{α} 亚基向着活性状态发生构象变化，此时 $G_{\beta\gamma}$ 不能再结合到 G_{α} 上。$G_{\beta\gamma}$ 并不发生构象的变化，但是当它被释放后，可以利用与 G_{α} 相同的方式结合到其他效应物蛋白上。

11.4.3 视觉系统

在包括膜结合受体在内的信号通路中，受体必须通过某种方式将胞外配体的结合信息转换成胞内的变化，这个变化又可以激活通路中的胞内其他部分。在 G 蛋白偶联受体这个例子中，通路需要 GEF 将三聚体 G 蛋白激活，而 GEF 在这个例子中就是 7 次跨膜螺旋受体本身。受体分子通常结合在特定的 G 蛋白上。受体的胞内卷曲部分必须结合在面向膜的 G 蛋白的某个部位上（图 11.24）。由于目前对受体和与之结合的 G 蛋白的蛋白质复合物结构尚未解析，所以受体识别特定 G 蛋白的机制还不清楚。

GDP 水解的机制到目前也没有研究清楚，但是可以推测该机制应当会包括受体胞质表面部分比较大的构象变化，而这种构象变化是由受体的一部分跨膜螺旋的相应移动造成的。跨膜螺旋的移动可能与之前我们提到的细菌视紫红质蛋白的构象变化是相似的。

受视导素调节的效应分子是 cGMP 磷酸二酯酶（PDE6），该分子在视杆细胞中与膜结合，由三个不同的亚基构成，即 α、β、γ 亚基。γ 亚基是该酶的抑制物，并且该亚基的一部分会与其他两个亚基结合来抑制酶的活性。活性状态下的视导素 α 亚基会与磷酸二酯酶的 γ 亚基羧基端相结合，从而去除 γ 亚基对该酶活性的抑制作用（图 11.25）。

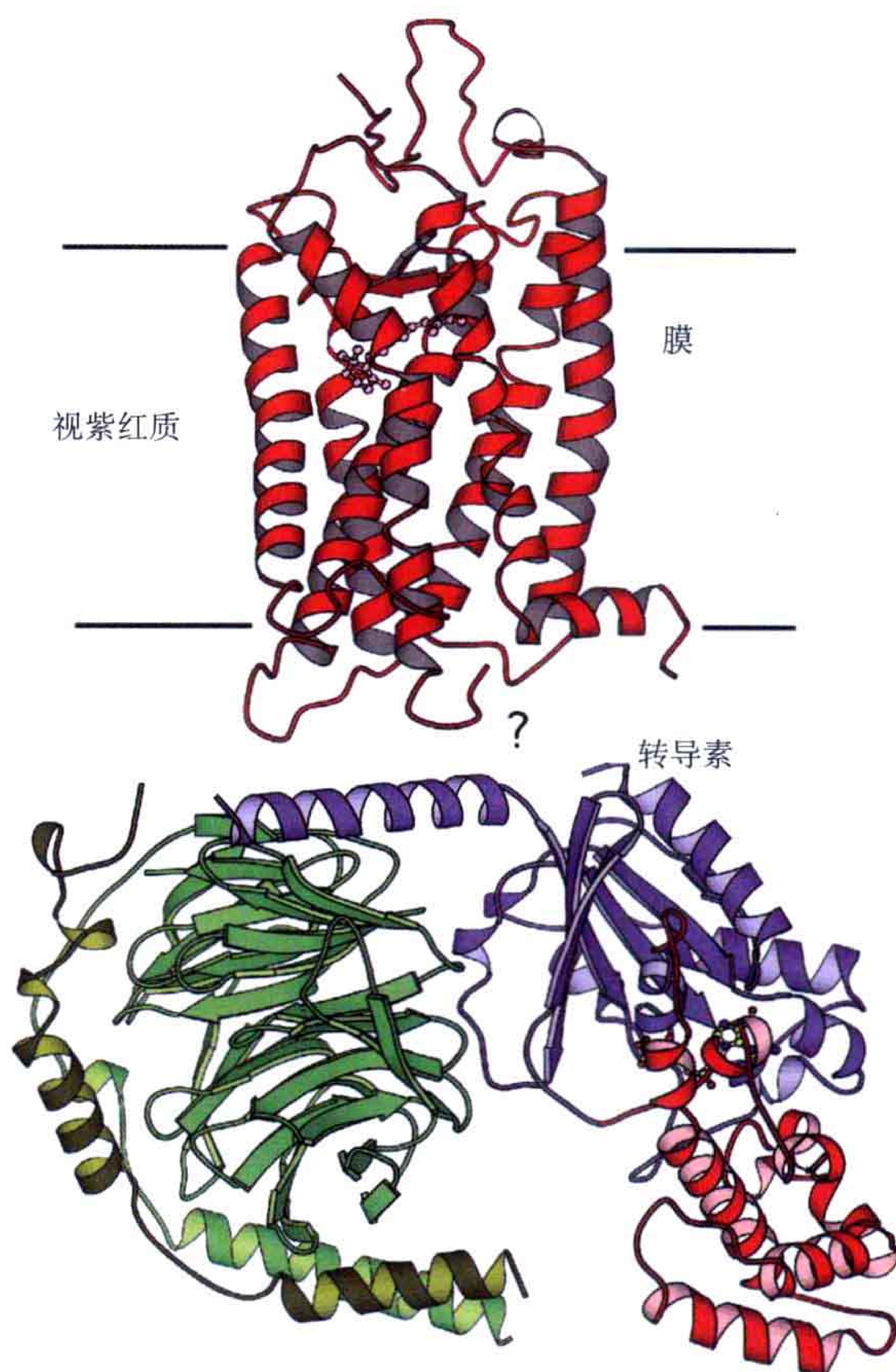


图 11.24 视紫质蛋白与 GDP 结合状态的视导素受体分子的推测排列情况。问号的标识是用来强调这两个分子间的相互作用细节并没有研究清楚。视导素分子利用疏水的 α 亚基氨基端与 γ 亚基羧基端结合到膜上。

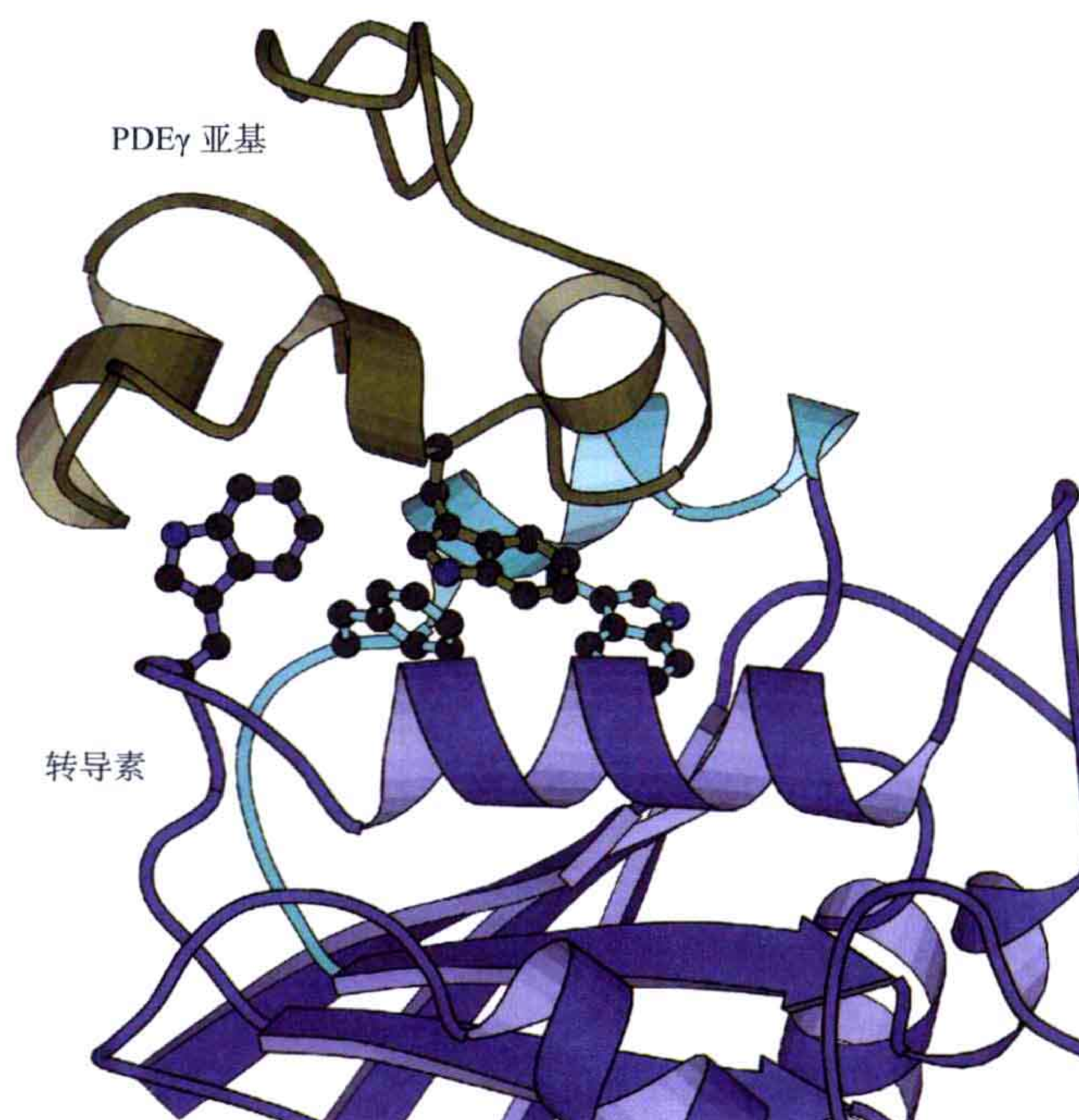


图 11.25 转导素（transducin）分子结合到 cGMP 磷酸二酯酶的 γ 亚基 (PDE γ subunit) 上（PDB：1FQJ）。switch Ⅱ卷曲和螺旋部分以蓝绿色标识。磷酸二酯酶的色氨酸残基与转导素 α 亚基 switch Ⅱ上的保守氨基酸相互作用。

在转导素与 cGMP 磷酸二酯酶 γ 亚基的复合物结构中，效应物分子上的一个芳香氨基酸插入 G 蛋白上 switch Ⅱ螺旋和 α3 螺旋间，与其上的保守氨基酸相互作用。该相互作用取决于 G 蛋白中核苷酸的状态，这是因为在 GDP 结合的蛋白质结构中，switch Ⅱ螺旋的方向是不同的。

11.4.4 腺苷酸环化酶的激活

G 蛋白偶联受体通路中主要的第二信使是环化单磷酸腺苷酸（cAMP）。该分子是腺苷酸环化酶的反应产物。人类腺苷酸环化酶含有 9 个不同的亚型。该酶可以催化 ATP 生成 cAMP，并受到多个信号通路的调控。GPCR 通路中激活研究得最清楚的效应物是腺苷酸环化酶，该酶受 G 蛋白 $G_{s\alpha}$ 调控而激活。

这个通路是以大量受体的激活为开始标识的，其中包括 β 肾上腺素受体的激活。这些受体都会作用于三聚体 G_s 蛋白，这里“s”代表刺激（stimulatory）。该酶较低的本底活性受到通过 G_s 复合物中 $G_{s\alpha}'$ 亚基调节而得到增强。腺苷酸环化酶其他亚型的激活分子工作机理不尽相同，这便使得在不同类型的细胞中，通过表达特定腺苷酸环化酶的亚基，使得这些细胞可以做出不同的反应。例如，尽管 $G_{i\alpha}$ 与 $G_{s\alpha}$ 十分相似，$G_{i\alpha}$ 却可以抑制某些腺苷酸环化酶的活性。

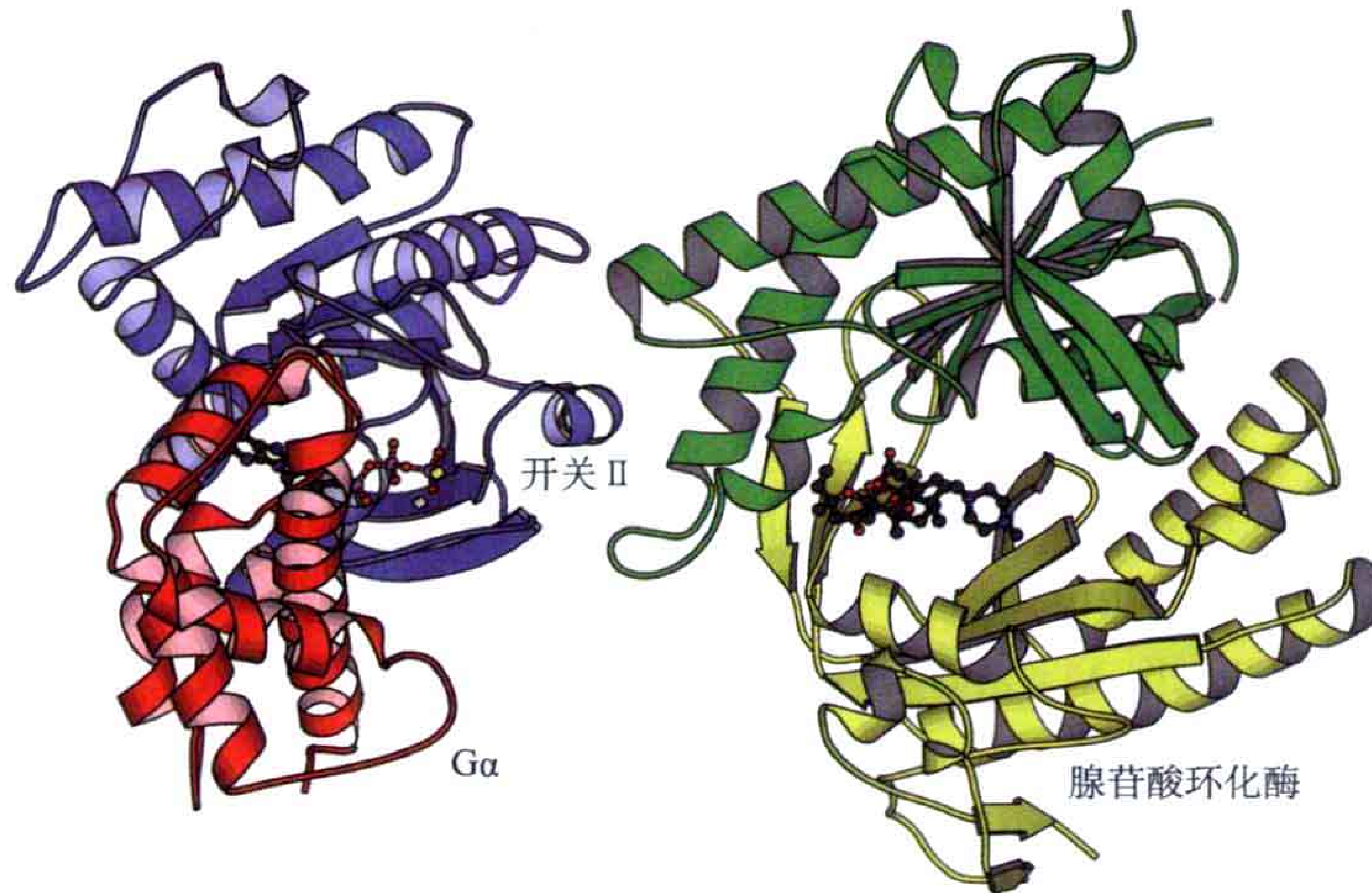


图 11.26 $G_{s\alpha}$ 亚基与腺苷酸环化酶催化部分的复合物结构。酶的两个亚基以黄色和绿色显示。腺苷酸环化酶催化部分的结合表面只能接受 GTP 结合后的蛋白质构象并将酶激活。一个 GTP 的底物类似物结合在 $G_{s\alpha}$ 亚基上。一个可以激活腺苷酸环化酶的小分子 forskolin 结合在酶上。

单个肽链构成的腺苷酸环化酶含有一个氨基端胞质部分和紧随其后的两个相似的结构单元（module）。每一个结构单元含有一个由 6 个跨膜螺旋组成的膜结构域和一个胞质催化部分。位于这两个结构单元上的两个胞质部分分别称为 C_1 和 $C_{2'}$ ，它们都含有两个结构域。其中的一个结构域 C_{1a} 和 $C_{2a'}$ 是同源的，并都处于催化活性状态，且体外重组的结构域依然具备催化活性。每个胞质部分的两个结构域形成异质二聚体（heterodimer）。C_{1a} 和 C_{2a} 的构象非常相似，它们都有 4 个反向平行折叠片层位于中间，两边夹有螺旋结构。折叠片层中的两个折叠束向外延伸形成了一个臂状亚结构域，羧基端也有一个由螺旋构成的亚结构域。折叠片层的拓扑结构与 DNA/RNA 聚合酶相似，ATP 的结合位点与聚合酶的 ATP 结合位点是相同的，两个保守（conserved）的天冬氨酸结合两个镁离子，这两个镁离子反过来又和 ATP 的磷酸相结合。酶反应活性中心位于两个结构域之间的作用面上，两个结构域都贡献了自己的一部分参与酶活反应。C_{1a} 和 C_{2a} 结构域，以及 $G_{s\alpha}$ 和另外一个效应物 forskolin 的复合物晶体结构已经得到解析（图 11.26）。$G_{s\alpha}$ 的 switch Ⅱ 螺旋结合在环化酶的沟槽中。GDP 结合构象中，这个螺旋所处的位置使得其不能再与环化酶相结合。这也是为什么只有 GTP 结合状态下的活性 G 蛋白才可以和活性的腺苷酸环化酶结合的原因。

延伸阅读

原始文献

Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D., and Kuriyan, J. (1998). The structural basis of the activation of Ras by Sos. *Nature* **394**, 337-343.

de Vos, A. M., Ultsch, M., and Kossiakoff, A. A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306-313.

Lambright, D. G., Noel, J. P., Hamm, H. E., and Sigler, P. B. (1994). Structural determinants for activation of the α-subunit of a heterotrimeric G protein. *Nature* **369**, 621-628.

Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E., and Sigler, P. B. (1996). The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* **379**, 311-319.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J. H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775-787.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-745.

Tesmer, J. J., Sunahara, R. K., Gilman, A. G., and Sprang, S. R. (1997). Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gs α.GTPγS. *Science* **278**, 1907-1916.

Xu, W., Harrison, S. C., and Eck, M. J. (1997). Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595-602.

综述文章

Hubbard, S.R. and Till, J.H. (2000). Protein tyrosine kinase structure and function. *Annu Rev Biochem*, **69**, 373-398.

Oldham, W.M. and Hammm H.E.. (2007). How do receptors activate G proteins? *Adv. Prot Chem*. **74**, 67-93.

Palczewski, K. (2006). G protein-coupled receptor rhodopsin. *Annu Rev Biochem*, **75**, 743-767.

Scheffzek, K. and Ahmadian, M.R. (2005). GTPase activating proteins: structural and functional insights 18 years after discovery. *Cell Mol Life Sci*, **62**, 3014-3038.

（范雪新 译　苏晓东　校）

第十二章

细胞的运动及物质运输

在真核细胞中，有许多用以维持细胞正确形状及运动的系统。这是一个由纤维组成的系统，即细胞骨架（cytoskeleton）。在细胞中同样存在另一个系统，负责将物质从内质网或者细胞膜上运输到细胞的其他部位。这个用于运输蛋白及其他分子的交通系统是由沿着纤维移动的囊泡组成的。另外，在许多动物的肌细胞中，存在一种纤维组成的分子，它可以使组织进行受控的运动。以上这些系统中的某些，在细菌中也能找到相似却更简单一些的版本。

12.1 肌动蛋白丝

12.1.1 肌动蛋白：一个可以组成动态纤维的蛋白质

肌动蛋白是肌肉纤维的主要组成部分之一，而肌肉纤维是属于中间纤维的一种细胞骨架。肌动蛋白同时也是细胞骨架中另一个成员——微管的主要组成部分。肌动蛋白的分子质量大约为 40kDa。

肌动蛋白有两种形式，即 G- 肌动蛋白和 F- 肌动蛋白。F- 肌动蛋白是肌动蛋白的纤维形式，在此情况下肌动蛋白分子形成长的螺旋。G- 肌动蛋白［G 指“球状的”（globular）］是单体形式的肌动蛋白。这两种形式互相之间存在一种平衡关系：G- 肌动蛋白在高浓度时会自发地转变为 F- 肌动蛋白，而在低浓度时肌动蛋白纤维便会解离。F- 肌动蛋白是有极性的，即肌动蛋白多聚体的两个自由末端是不同的。F- 肌动蛋白的生长主要发生在一个方向，即“+”方向。这个末端又被称为是“倒钩（barbed）”末端，而另一个末端则被称为“尖（pointed）”末端。在中间浓度时，F- 肌动蛋白可以在“+”末端生长，同时在“−”末端解离，这个过程被称为踏车运动（tread-milling）。

肌动蛋白可与 ATP 结合，并且这个结合对于肌动蛋白纤维的生长是十分重要的。肌动蛋白必须与 ATP 结合后才倾向于结合到纤维的“+”末端。当 ATP 与肌动蛋白纤维结合后，肌动蛋白中的 ATP 被降解为 ADP。这意味着结合 ATP 的肌动蛋白被添加到“+”末端，而在“−”末端失去的则是结合 ADP 的肌动蛋白。肌动蛋白中 ATP 的水解在某

种程度上类似 G- 蛋白中 GTP 到 GDP 的转变：这是一种纤维降解的定时装置。

肌动蛋白的多聚化对于许多细胞特性都是重要的，如细胞形状和运动。因此，在细胞中 F- 肌动蛋白的生长受到许多蛋白质的调控，有一些负责阻止多聚化和去多聚化发生（capping protein，帽蛋白），而另一些则可以把肌动蛋白纤维切成更小的片段（severing protein，剪切蛋白）。还有许多蛋白质参与把肌肉及细胞骨架中的肌动蛋白纤维交联起来。这些蛋白中的某些蛋白参与肌动蛋白纤维束的形成，其他一些蛋白质则将肌动蛋白纤维组合成三维的交联网络。

在人类体内，共有 6 种不同的肌动蛋白基因，这些基因不同的表达情况取决于细胞的类型。α 肌动蛋白存在于肌肉组织中。而 β 和 γ 肌动蛋白则同时存在于大多数的细胞类型中，它们是细胞骨架的组成部分，并参与了细胞内部的运动。

当比较不同类型肌动蛋白的序列时，可以发现仅存在一些很小的变化。肌动蛋白是最保守的蛋白质之一。对于肌动蛋白及其他一些高度保守的蛋白质来说，其蛋白质表面是很重要的，因为所有与其他蛋白的特异性相互作用都是发生在蛋白质表面的，因此，这些蛋白质的氨基酸序列在长期的进化过程中都保持着很少的变化。

12.1.1.1 肌动蛋白单体

单体肌动蛋白是一个相对来说比较紧密的蛋白质。通常描述肌动蛋白的结构时，可以描述为它有两个结构域，而每一个结构域则又可以被分为两个亚结构域（图 12.1）。而更精确的描述是，肌动蛋白单体可以划分为 4 个结构域，其中结构域 2 是结构域 1 的插入结构域（insertion），而结构域 4 是结构域 3 的插入结构域。C 端的 40 个残基形成了一些螺旋结构，这些螺旋结构是结构域 1 的一部分，而这 40 个残基在序列上则是位于结构域 3 的后面。ATP 所结合的活性位点在一个很深的裂隙之中，这个裂隙位于分子的中间。结构域 1 和 3 有着同样的拓扑结构。这些结构域的主要由中心的平行的折叠片层及两侧的螺旋所组成。同样的折叠方式也出现在一些看似与肌动蛋白无关的蛋白质之中，如分子伴侣 hsp70、己糖激酶及其他一些激酶。虽然这些蛋白质的功能是各异的，但有一个共同点是其活性需要 ATP 参与。

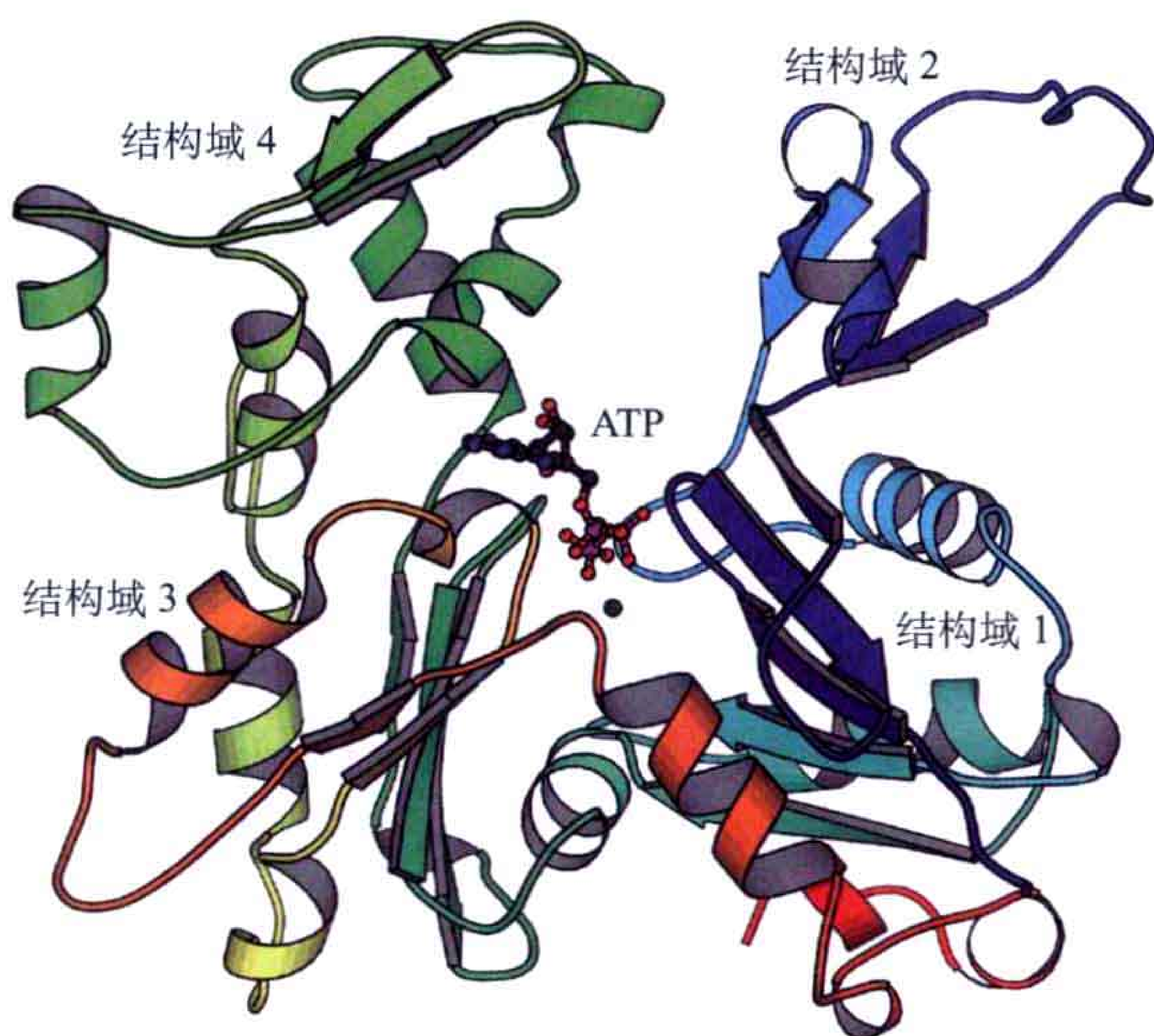


图 12.1 肌动蛋白单体的示意图。“–”末端位于上方。图中的颜色指示从 N 端（蓝色）至 C 端（红色）（PDB：1ATN）。Domain1~4：结构域 1~4。

己糖激酶分子是一个关于结构域转动的著名例子，它的结构域转动是由结合它的底物葡萄糖所引起的。因此人们相信由于与己糖激酶的相似性，肌动蛋

白也应该可以发生涉及结构域转动的构象变化。一些不同形式的肌动蛋白的晶体结构显示出微小的结构域转动，但这比在己糖激酶中所观察到的转动的量级要小得多。由于结合 ADP 和结合 ATP 的肌动蛋白有着与肌动蛋白纤维不同的亲合性，将这种亲合性差异的原因解释为结构域转动是很合理的。但是，这两个形式的肌动蛋白晶体结构十分相似，且没有结构域转动（图 12.2）。

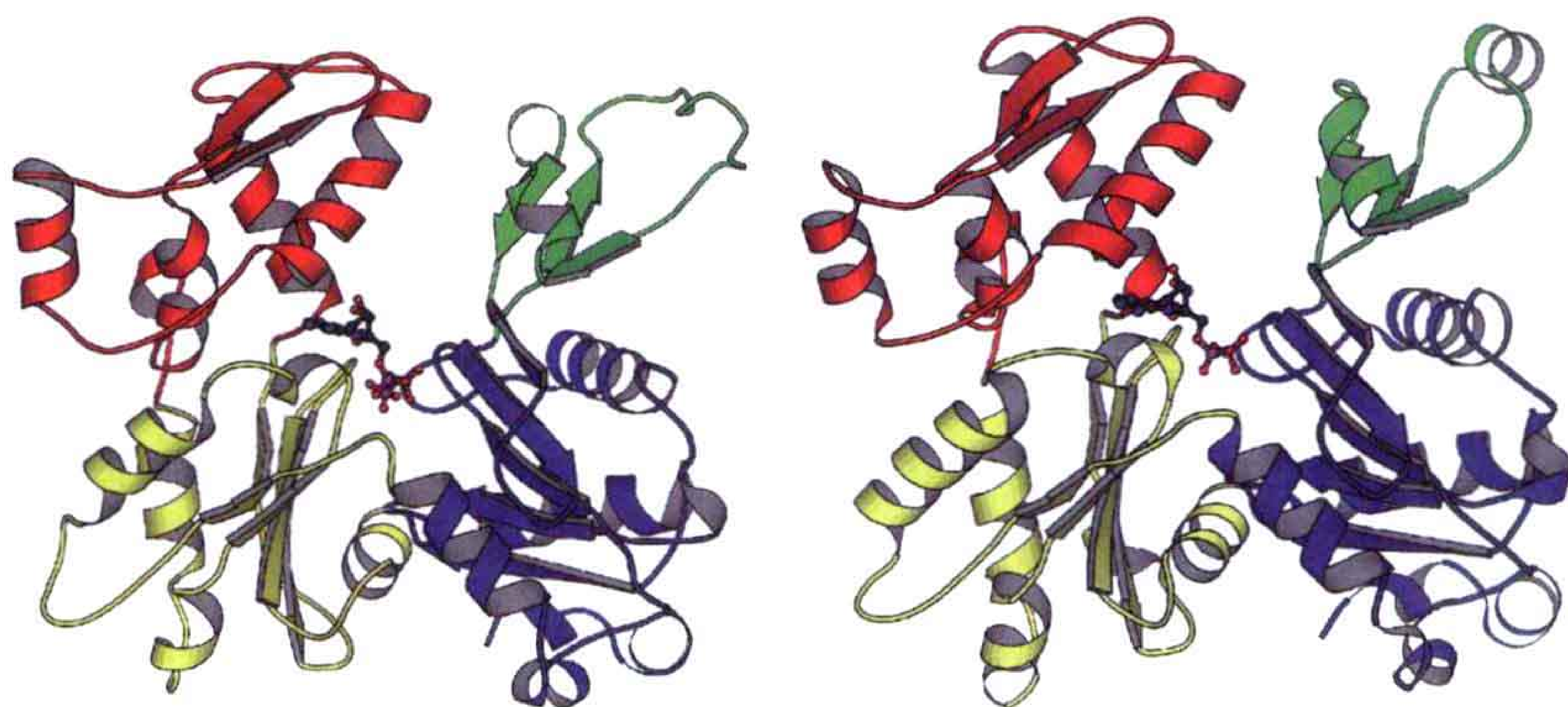


图 12.2　结合 ATP（左）与 ADP（右）的肌动蛋白构象的比较（PDB：1J6Z）。在亚结构域 2（绿色）中有一些结构重排，但是在整个晶体结构中并没有发现大的结构域重新排列的现象。

12.1.1.2　肌动蛋白纤维的结构

肌动蛋白纤维的结构可以通过纤维衍射来研究。虽然纤维衍射得到的电子密度的分辨率十分有限，但是通过将已知结构的单体肌动蛋白放入得到的电子密度图中，可以得到合理的肌动蛋白纤维结构模型。肌动蛋白纤维可以看作由两股螺旋所组成的超螺旋（图 12.3）。结构域 2 是肌动蛋白分子中暴露在最表面的部分。单体肌动蛋白不但与沿着螺旋的相邻分子产生相互作用，超螺旋的两股“螺旋股”之间也会有相互作用。

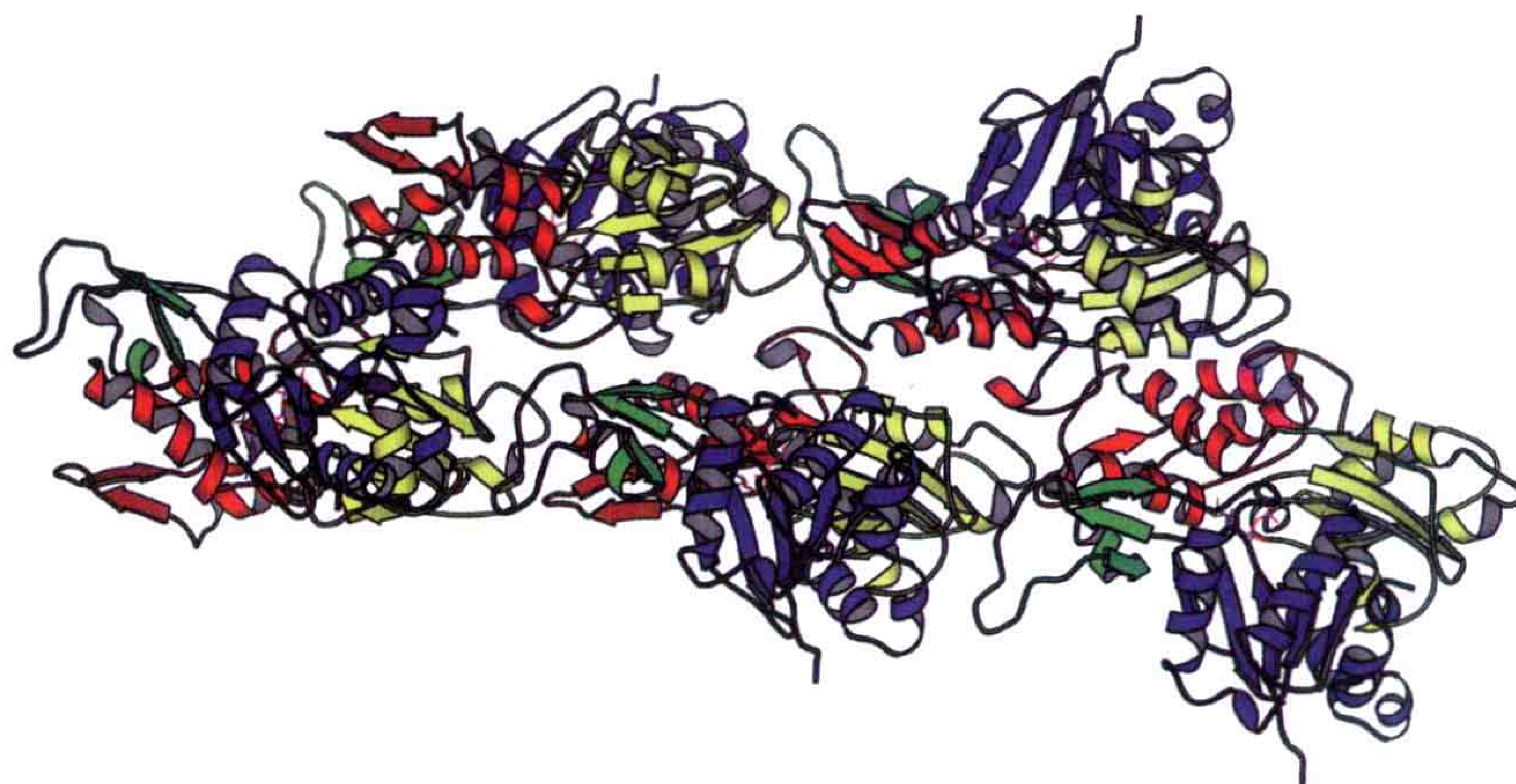


图 12.3　肌动蛋白纤维的结构模型。图中所示为 5 个肌动蛋白单体，在肌动蛋白的双螺旋结构中，三个单体属于下方的螺旋股而另两个单体属于上方的螺旋股。“–”末端（即图 12.1 的上方）在本图中位于左边。结构域的颜色标注方式与图 12.2 相同（PDB：1ALM）。

单体肌动蛋白的 ATP 水解酶活性很低，而在形成纤维后被激发。这个激发的机制目前并不清楚，但应该与 G 蛋白中的一些 GAP 蛋白相似，即通过与其他分子相互作用，适合酶催化形式的构象被稳定了。

12.1.2 肌动蛋白多聚化的调控

12.1.2.1 与单体肌动蛋白结合的蛋白质

有许多蛋白质都是与肌动蛋白结合的。其中一些蛋白质的主要功能是调控肌动蛋白的多聚化。这些蛋白质有的与单体肌动蛋白结合，参与阻碍肌动蛋白纤维的生长（加帽），另一些则是负责剪切肌动蛋白纤维（表 12.1）。

表 12.1 参与肌动蛋白多聚化调控的蛋白质举例

名称	功能	与肌动蛋白结合的结构域的结构
胸腺素 β_4（thymosin β_4）	储存单体的肌动蛋白	非球状的折叠，其螺旋与肌动蛋白的不同位置结合
肌动蛋白抑制蛋白 (profilin)	结合单体肌动蛋白，核苷酸交换	肌动蛋白抑制蛋白结构域
凝溶胶蛋白（gelsolin）	肌动蛋白纤维“+”末端加帽，剪切	凝溶胶蛋白结构域
Arp2/3 复合体	与“–”末端结合并且在肌动蛋白纤维的侧面形成新的“+”末端	与肌动蛋白类似

12.1.2.2 胸腺素 β_4

一个被称为胸腺素 β_4 的小蛋白可以在细胞内结合肌动蛋白。这个蛋白质主要负责维持单体肌动蛋白的高数量。它在溶液状态下并没有一个紧密的球状结构。当结合肌动蛋白时，它的 N 端螺旋与肌动蛋白的“+”末端结合，同时另外的螺旋与“–”末端结合。一个延伸的片段连接着这些螺旋。通过这种方式，这个小蛋白可以同时包裹肌动蛋白单体的两个末端。

许多其他参与调控肌动蛋白聚集的蛋白质都含有一个小的结构模块，这个模块被称为 WH2（WASP homology 2）。这个模块与胸腺素蛋白的 N 端有着相似的结构，并且形成一个螺旋用于与肌动蛋白的一个凹槽结合，这个凹槽位于肌动蛋白结构域 1 和 3 之间（图 12.4）。

12.1.2.3 肌动蛋白抑制蛋白

肌动蛋白抑制蛋白是一个被广泛研究的肌动蛋白结合蛋白，它参与肌动蛋白纤维的稳定及生长。它的生化功能之一似乎是使 ADP- 肌动蛋白重新变成 ATP- 肌动蛋白，同时它仍然参与其他类型的细胞骨架调控。

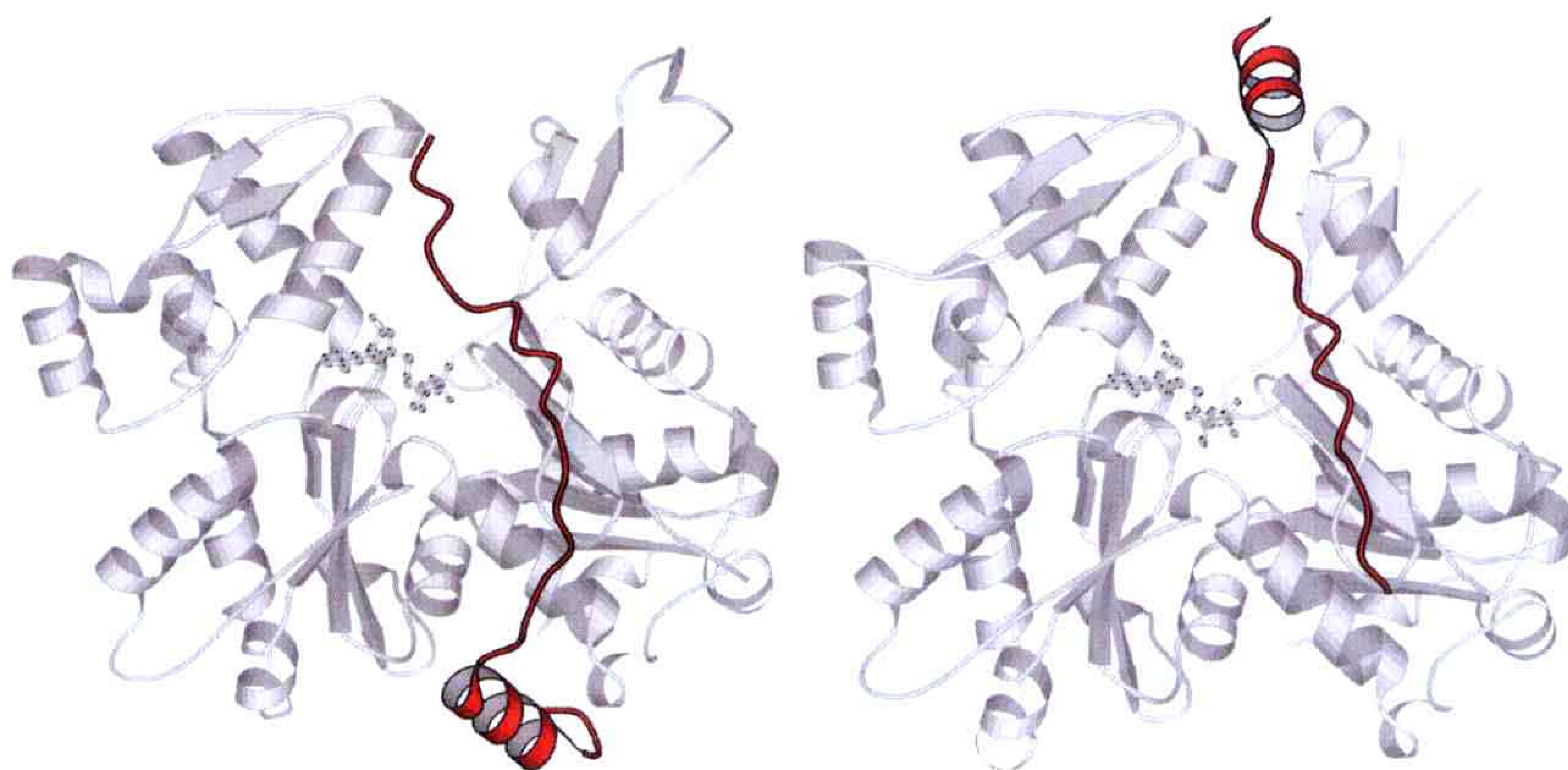


图 12.4　阻止单体肌动蛋白的多聚化。左：肌动蛋白与一个 WH2 结构模块（红色）的结合。这个 WH2 结构域是 WASP- 相互作用蛋白的一部分（PDB：2A41）。一个 N 端的螺旋结合在肌动蛋白的“+”末端，并且结合在肌动蛋白结构域 1 和结构域 3 之间的沟槽中。而 WH2 模块的一个延伸部分则沿着整个肌动蛋白分子结合。右：胸腺素 β_4 的 C 端（红色）与肌动蛋白的“–”末端结合（PDB：1T44）。在晶体结构中，这个片段连接着凝胶蛋白的一个结构域，这个凝胶蛋白结合在肌动蛋白的“+”末端。这两个片段结合在一起被认为可以代表胸腺素 β_4 与肌动蛋白的结合。

肌动蛋白与肌动蛋白抑制蛋白的复合体结构显示，肌动蛋白抑制蛋白结合在肌动蛋白远离活性位点的一侧，这解释了为何 ADP 可以解离（图 12.5）。而这也解释了为何肌动蛋白与肌动蛋白抑制蛋白复合体只能被加在肌动蛋白纤维的球状（“+”）末端。

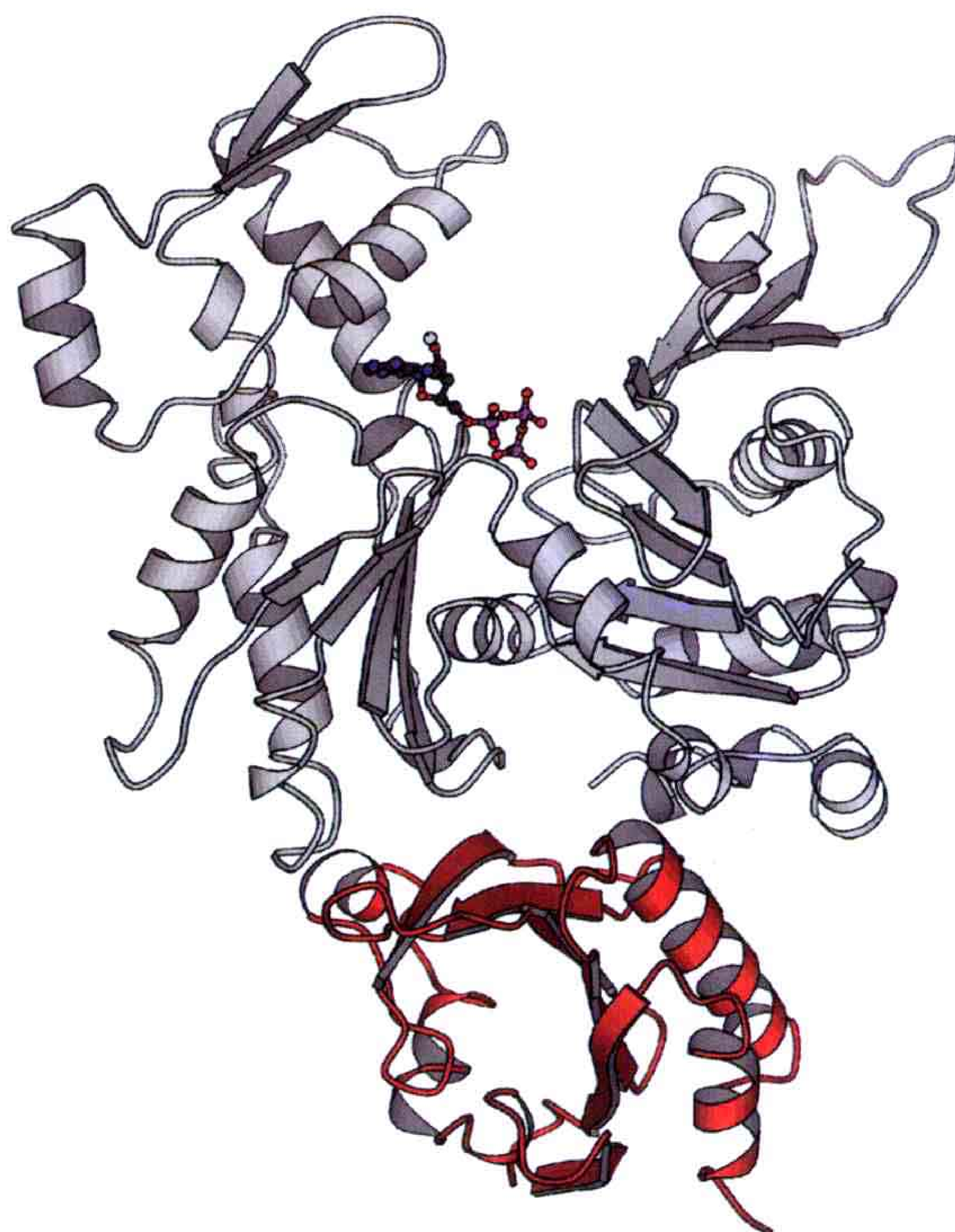


图 12.5　肌动蛋白与肌动蛋白抑制蛋白复合体的示意图（PDB：1HLU）。肌动蛋白的“+”末端被肌动蛋白抑制蛋白（红色）所封闭。

12.1.3 肌动蛋白纤维的加帽或剪切蛋白

12.1.3.1 凝溶胶蛋白

凝溶胶蛋白是一种在真核细胞中可以剪切肌动蛋白纤维或者为肌动蛋白纤维加帽的蛋白质。它其中一个生理功能是阻止血浆中的肌动蛋白形成纤维，同时它还可以在细胞内调节细胞骨架中的肌动蛋白纤维。一个蛋白质同时用于细胞内和细胞外，这是不多见的。凝溶胶蛋白还可以在纤维形成过程中参与成核。

凝溶胶蛋白含有 6 个构象相似的结构域。每一个结构域都有一个混合的片层，一些螺旋位于片层的一侧。在完整的分子中，结构域 1 ～ 3 与结构域 4 ～ 6 是整体相似的。这暗示这个分子在进化上先经历了一次祖先基因的三倍化（triplication），接下来又经历了一次这个三结构域基因的双倍化（duplication）。结构域 1 和 3 形成了一个连续的片层，而结构域 4 和 6 也是如此。在调控肌动蛋白的多聚化时，凝溶胶蛋白的不同结构域有着不同的功能。结构域 1 主要负责剪切肌动蛋白纤维。

已知的晶体结构有：凝溶胶蛋白的一个单独结构域与肌动蛋白的复合体、凝溶胶蛋白的结构域 1 ～ 3 与肌动蛋白的复合体，以及凝溶胶蛋白结构域 4 ～ 6 与肌动蛋白的复合体。在复合体中，凝溶胶蛋白的结构域 1 或 4 与肌动蛋白“+”末端的结合方式与肌动蛋白抑制蛋白相同。一个螺旋结合在 WH2 结构域结合肌动蛋白的相同位置（图 12.4）。结构域 1 ～ 3 或结构域 4 ～ 6 与肌动蛋白的复合体结构显示出一个大的结构域重排。结构域 1 和 3 之间及结构域 4 和 6 之间形成的连续的片层被破坏了，从而释放出了结合肌动蛋白的表面。在结构域 1 和 4 中，一个天冬氨酸侧链分别与结构域 3 和 6 中的赖氨酸侧链相互作用，以稳定未结合状态的凝溶胶蛋白结构。而在结合了钙离子后，这个天冬氨酸侧链与钙离子结合，从而打破了原本的盐键，于是结构域间的连接被削弱，使凝溶胶蛋白的肌动蛋白结合表面得以与肌动蛋白结合（图 12.6 和图 12.7）。

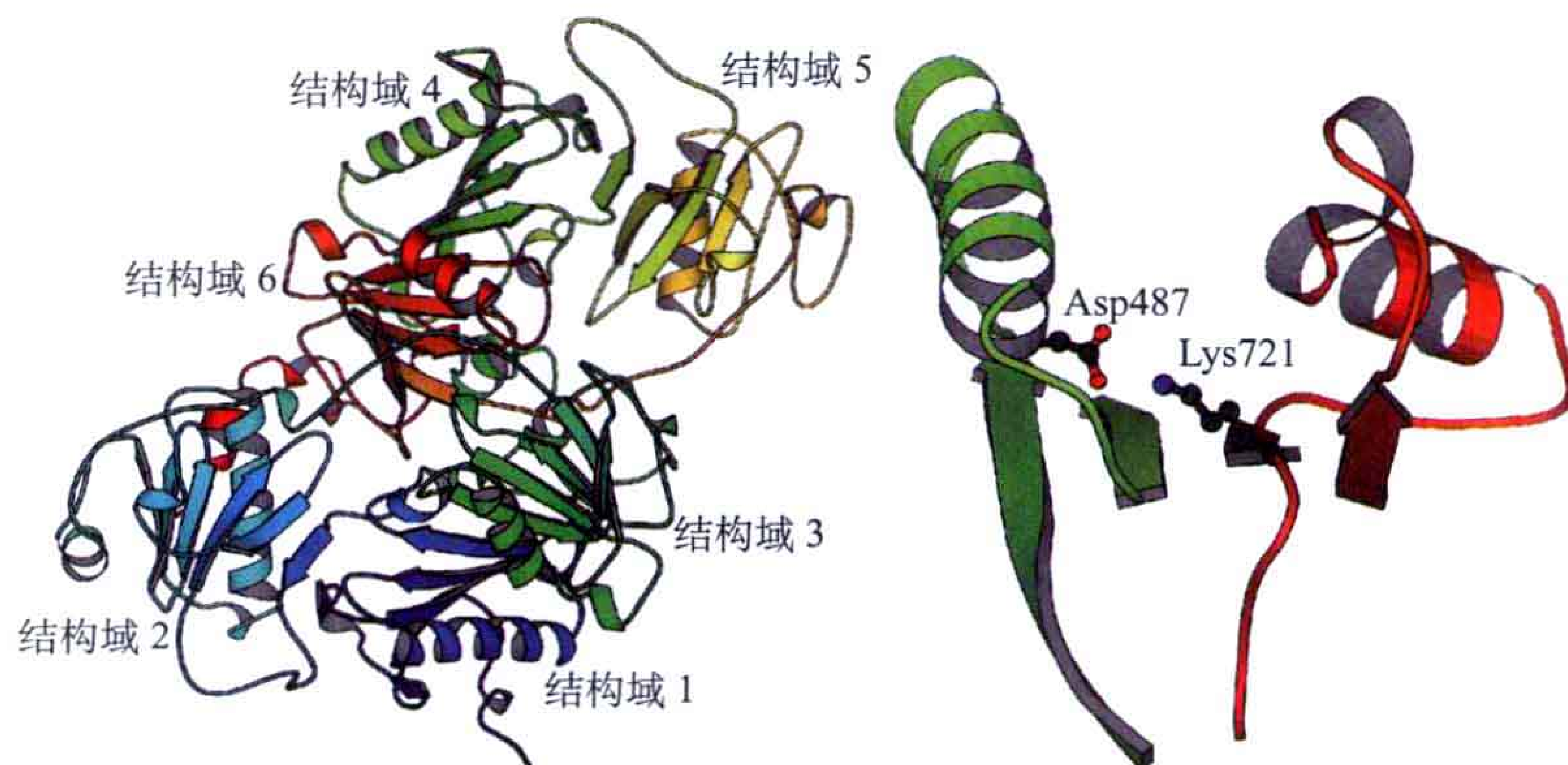


图 12.6　血浆中凝溶胶蛋白的结构。在这个构象中，肌动蛋白结合表面被完全覆盖了。通过钙离子的调节，复合体中相应的连接被破坏，从而把肌动蛋白结合表面给暴露出来。右：结构域 4 和 6 相互作用的细节图，即 487 位的天冬氨酸（结构域 4）和 721 位的赖氨酸（结构域 6）形成了盐键，用以稳定结构域间的连接（PDB：1D0N）。

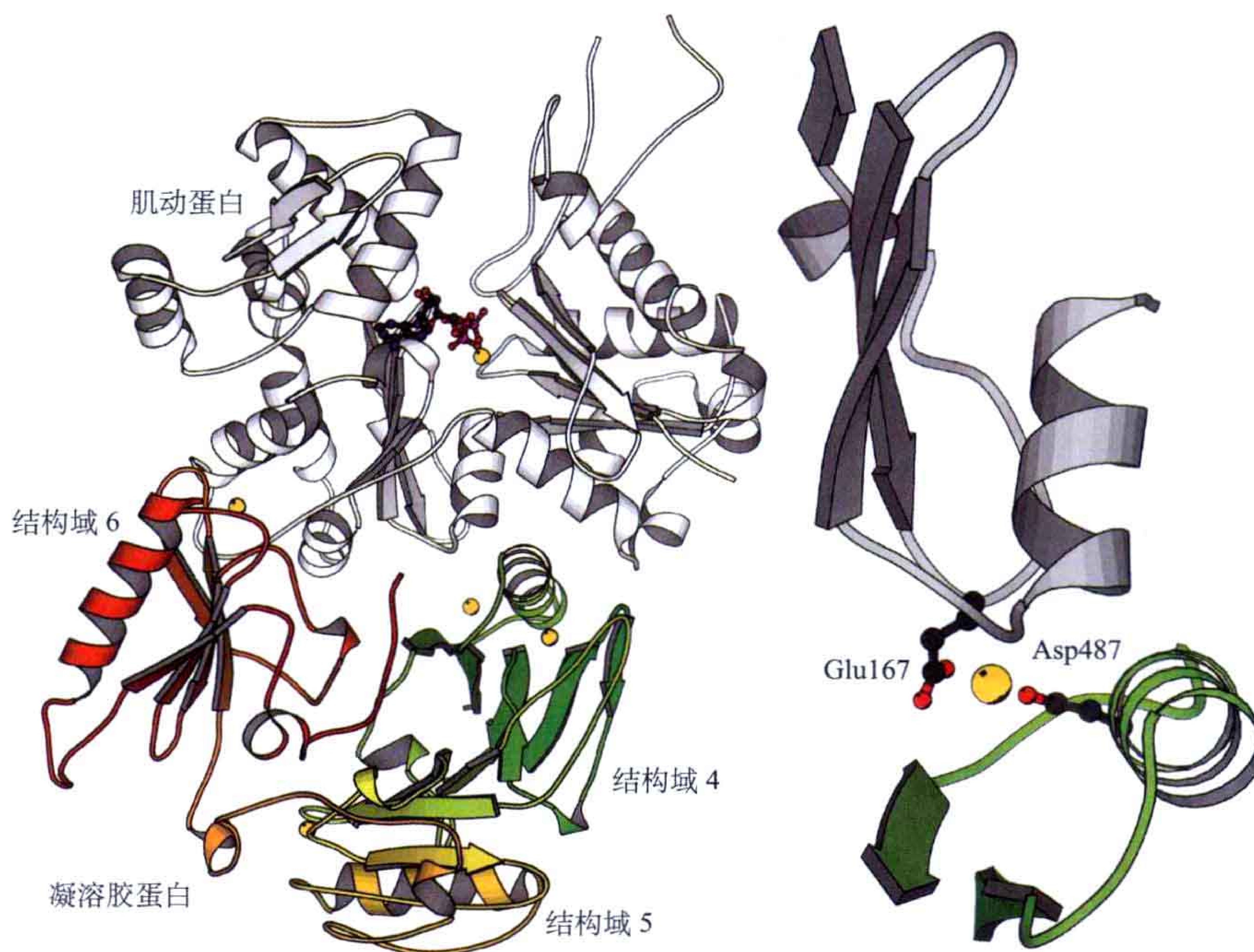


图 12.7　肌动蛋白（灰色）与凝溶胶蛋白结构域 4~6 的复合体结构。我们注意到在复合体结构中，结构域 4 和 6 之间连续的片层被打破，这很可能是由于钙离子的活化。右：结合表面的细节图。其中凝溶胶蛋白结构域 4 的 487 位天冬氨酸和肌动蛋白的 167 位谷氨酸都与钙离子相结合（PDB：1H1V）。

12.1.4　肌动蛋白的附着及交联

12.1.4.1　Arp2/3 复合体

在有运动能力的真核细胞中，细胞的前沿是被生长的肌动蛋白网络所推动的（图 12.8）。新的肌动蛋白纤维像树枝一样从其他纤维中分岔而生长出来。许多蛋白质参与调控这些肌动蛋白纤维的生长和交联。一个被称为 Arp2/3 的蛋白复合体负责在这个肌动蛋白网络中产生新的分岔。与此同时，另一些蛋白质，如切丝蛋白（cofilin）促使肌动蛋白网络中较“老”的分岔发生去多聚化。

Arp2/3 是由 7 个蛋白质形成的复合体。这个复合体的晶体结构已解析（图 12.9），通过此结构发现 Arp2 和 Arp3 链与肌动蛋白是十分相似的，但 Arp2 的 N 端结构域由于柔性过强而无法给出合适的模型。图 12.10 显示的是 Arp3 的结构。与肌动蛋白相比，Arp3 的一些 loop 稍长了一些。Arp2 和 Arp3 可能是新肌动蛋白纤维发生多聚化的起点。复合体的剩余部分用于与肌动蛋白纤维结合，并且起调控成核作用。这个复合体是没有活性的，直到与肌动蛋白纤维结合并且被其他蛋白激活后才有活性，如被 WASP 蛋白所激活。

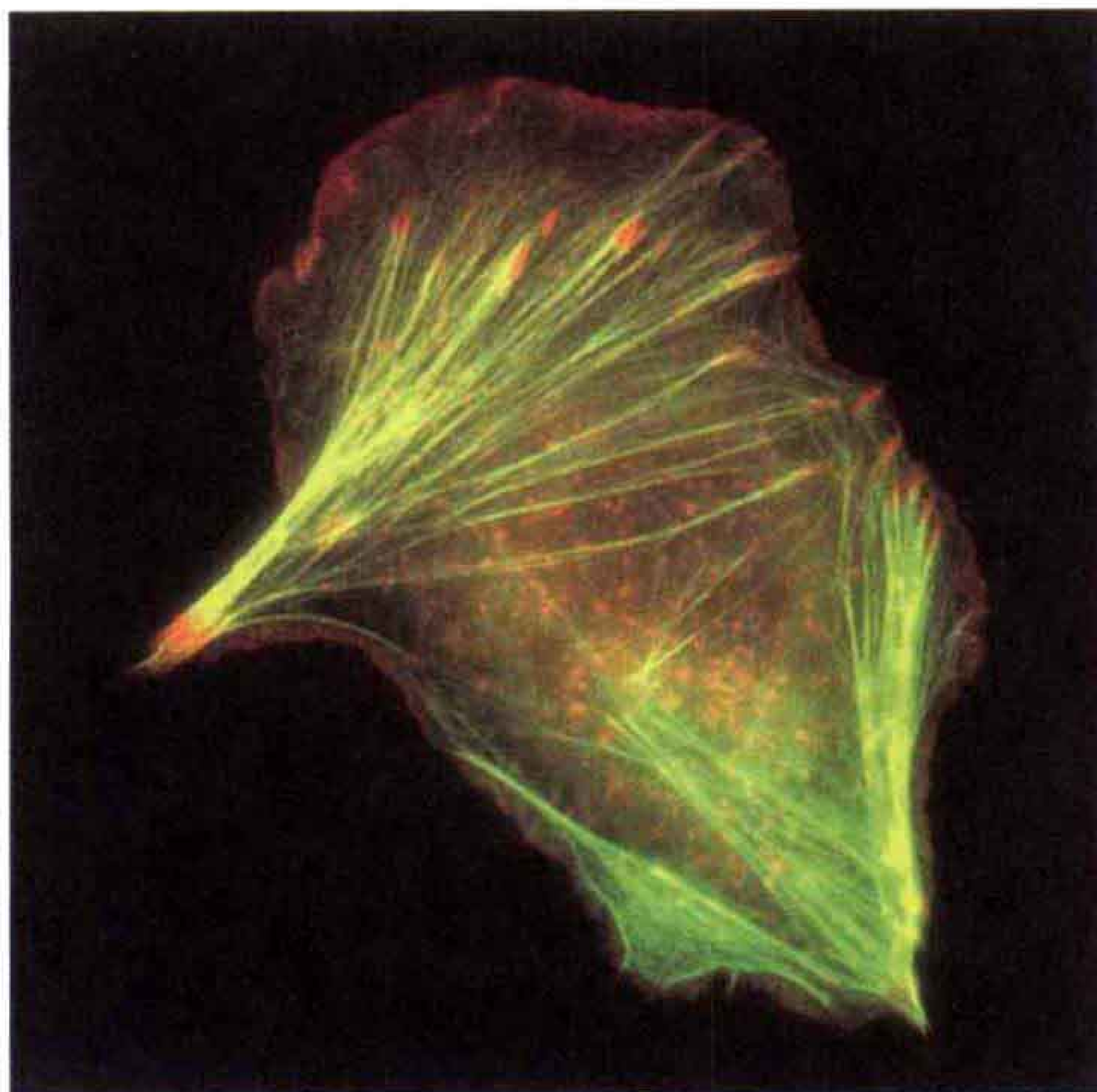


图 12.8 一个运动中的细胞。此图显示的是一个人工培养的肿瘤细胞（melanoma，黑色素瘤）。荧光显示的纤维（应力纤维）代表了一个属于中间纤维的蛋白质多聚体系统，这个系统是由肌动蛋白组成的，肌动蛋白参与了产生动力以完成细胞运动和迁移。数百个蛋白质分子组成了这个系统。这个细胞被转染了一个质粒，使得可以表达一种肌动蛋白的调控蛋白，即原肌球蛋白（tropomyosin）[TM5，带有绿色荧光蛋白（GFP）标签]。肌动蛋白是通过罗丹明 - 鬼笔环肽（rhodamine phalloidin，红色）显色的。根据 GFP-TM5（绿色）与肌动蛋白（红色）的平衡情况的不同，含有纤维结构的肌动蛋白的颜色从红色到黄色再到绿色不等。应力纤维末端的红色斑点表示那些位于没有原肌球蛋白的黏性位点的肌动蛋白。本图由 Louise Bertilsson 所拍摄，并由 Uno Lindberg 提供。

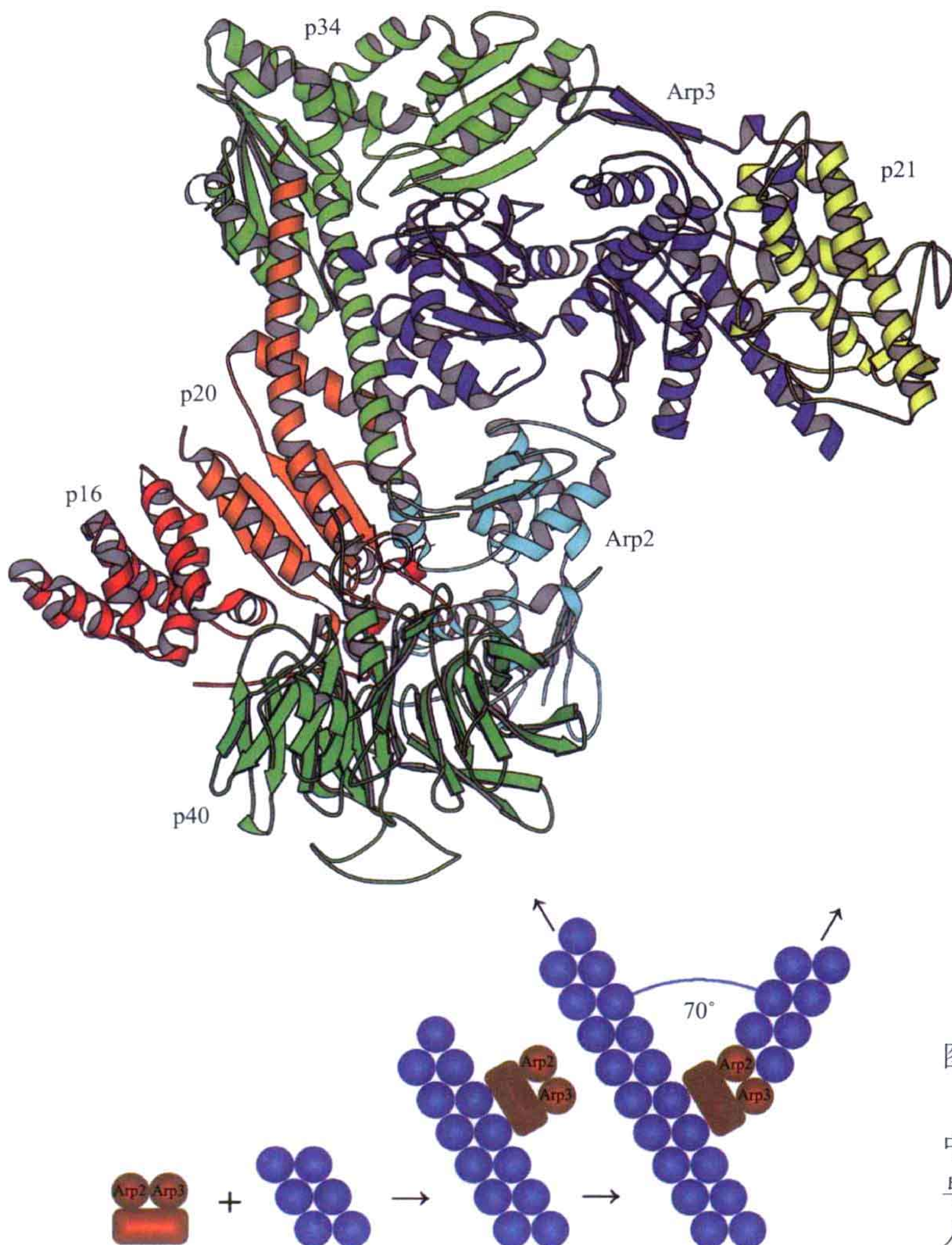


图 12.9 上：Arp2/3 复合体（PDB：1K8K）；下：设想中这个复合体（橙色）如何与肌动蛋白（蓝色）结合引入分岔的示意图。

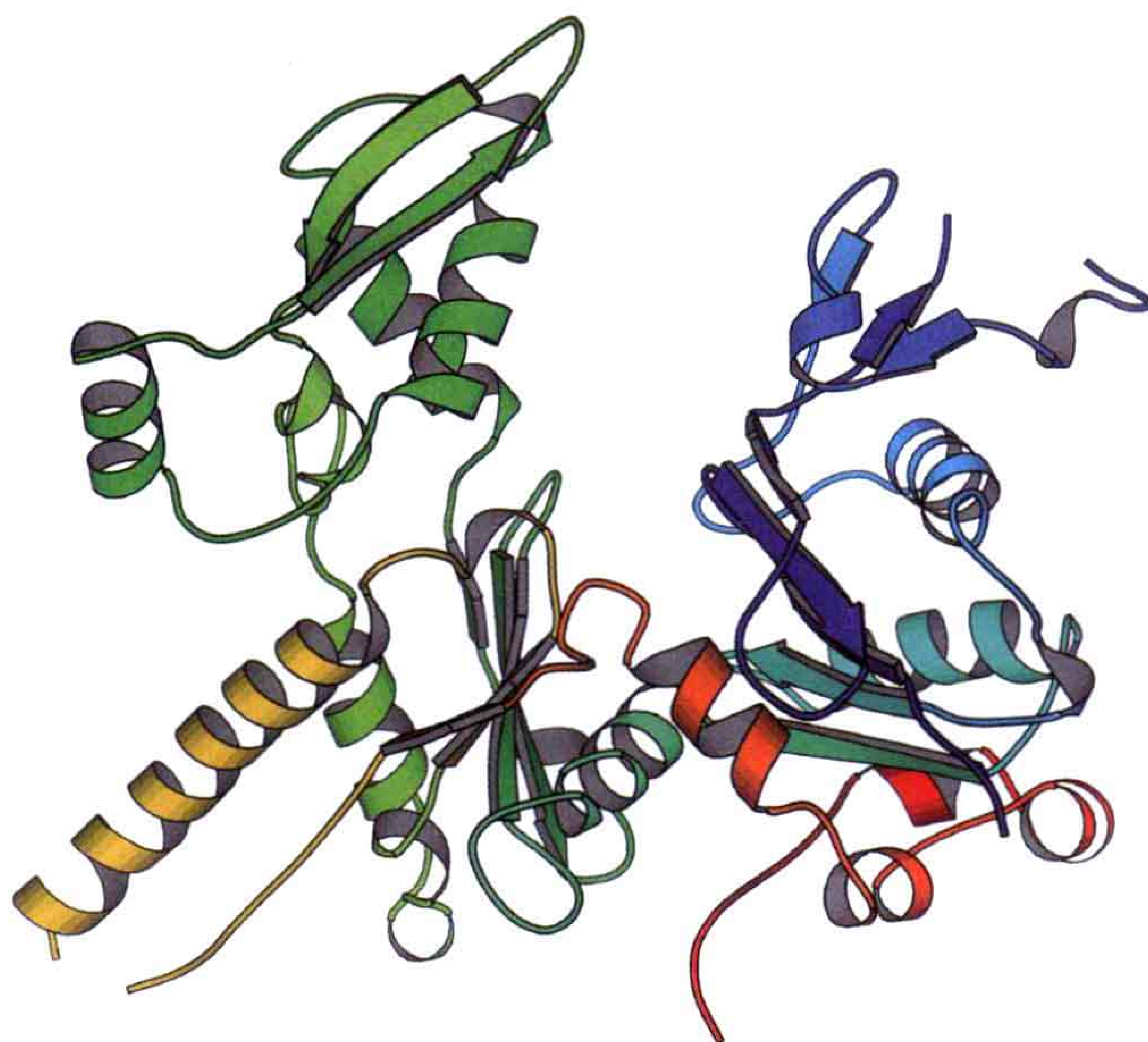


图 12.10　Arp2/3 复合体中类似肌动蛋白的 Arp3 亚基的示意图。在这个结构中，结合核苷酸的裂隙比肌动蛋白要打得更开一些。

另一个拥有常规折叠方式的结构域是螺旋桨式结构域 p40。这个结构域十分像三聚体 G 蛋白中参与细胞信号转导的 β 亚基（见 11.4 节）。在 Arp2/3 复合体剩下的结构中，p34 和 p20 有着相同的拓扑结构，即有一个反平行的片层，这个片层属于卷曲类型，并且在 C 端有一个长的螺旋。p34 有两个这样的折叠方式。p21 和 p16 蛋白则都是螺旋结构。

虽然 Arp2/3 复合体的结构为深入研究其与肌动蛋白纤维结合的机制及如何起始肌动蛋白多聚化提供了一个起点，但究竟多聚化如何开始仍然有待研究。最有可能的是，这个复合体在结合了肌动蛋白纤维或者被 WASP 蛋白激活后会经历一个构象变化。

12.2　肌球蛋白与肌肉功能

12.2.1　肌肉的结构

肌肉细胞中含有成束的收缩性蛋白纤维，这种纤维被称为肌原纤维（myofibril）。在肌原纤维中，我们可以看到（通过电子显微镜）重复的明暗相间的条纹及盘状结构（称为 Z 盘，Z disc，图 12.11）。暗的条纹中含有粗丝（thick filament）而明条纹中含有细丝（thin filament）。Z 盘是用于细丝附着的结构。肌原纤维上的一个重复单元被称为

肌小节（sarcomere），肌小节从一个 Z 盘开始到下一个 Z 盘结束。粗丝是由约 300 个肌球蛋白分子交联在 M 条带上形成的。细丝的主要组成部分是肌动蛋白纤维。细丝以一个规则的排列方式部分嵌入粗丝中（图 12.12）。除了肌球蛋白和肌动蛋白外，还有一些其他蛋白质与肌丝和 Z 盘有关。其中一个蛋白质称为肌联蛋白（titin），它是一个非常长的蛋白质，分子质量为 2.7 MDa。它从 M 条带延伸到 Z 盘并且似乎参与维持肌小节的长度及在肌小节中间保持 M 条带的位置。

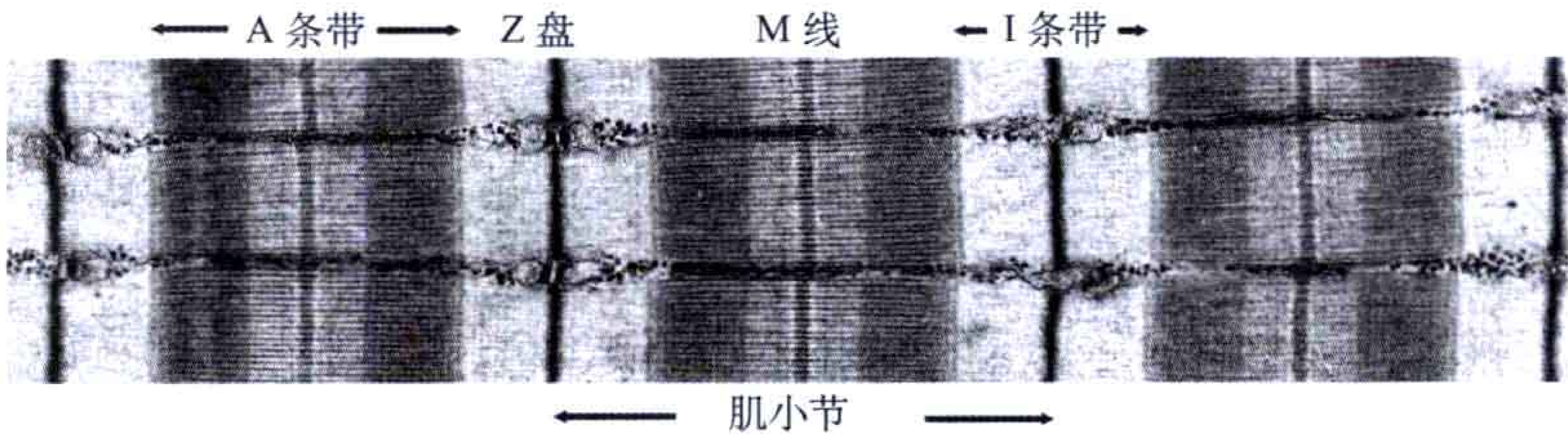


图 12.11　电镜下的肌肉纤维。暗的 A 条带是由粗丝组成的，而明亮的 I 条带是由连接在 Z 盘上的细丝组成的。粗丝被交联在 M 线上。最暗的区域是粗丝和细丝交叠处（此图由马萨诸塞大学的 Roger Craig 博士提供。）

肌肉的收缩运动是通过肌丝的滑动从而缩短肌小节而完成的。细丝和粗丝通过蛋白 - 蛋白相互作用的方式交联在一起。肌丝的滑动是通过粗丝和细丝间交联的形成与打破完成的，由此将粗丝拉得更靠近 Z 盘（图 12.12）。

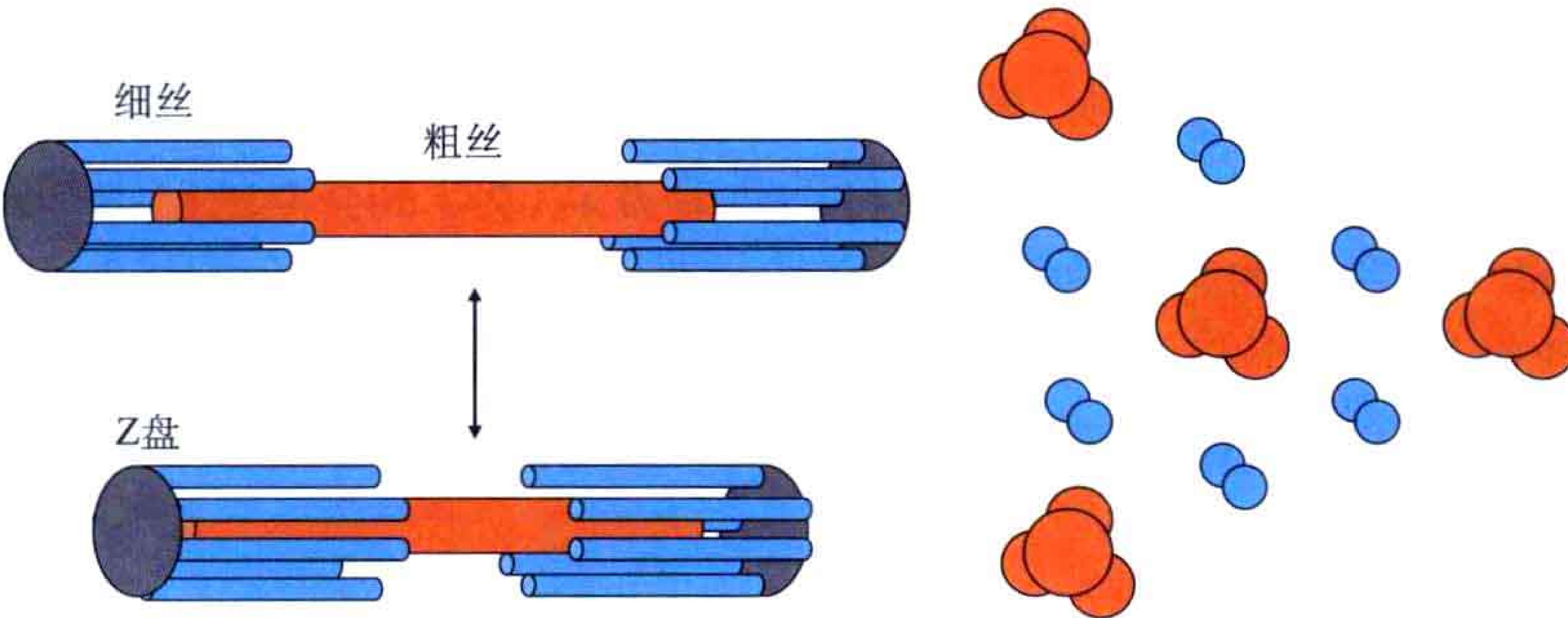


图 12.12　分别处于放松及收缩状态的肌小节的示意图。右图示意细丝（蓝色）和粗丝（橙色）在它们相互作用区域的排列方式。

12.2.2　与细丝结合的蛋白质

12.2.2.1　原肌球蛋白

细丝与两个蛋白质相互作用，这两个蛋白质参与了钙依赖的肌肉收缩调控。其中一个蛋白质被称为原肌球原白（tropomyosin），这是一个十分细长的蛋白质。原肌球蛋白是由两个相同的 α 螺旋链所组成，这两个螺旋链形成了一个卷曲螺旋（coiled-coil）

结构。它的序列中含有 40 个七氨基酸重复序列的片段，即 abcdefg，其中 a 和 d 绝大多数是非极性的氨基酸。这种排列方式是典型的卷曲螺旋蛋白所拥有的。这一对螺旋结构以头对尾的形式连接，形成了一个围绕着细丝中肌动蛋白螺旋的连续结构。每一个原肌球蛋白二聚体与 7 对肌动蛋白单体结合。

12.2.2.2　肌钙蛋白

另一个与细丝结合的分子被称为肌钙蛋白（troponin），这是一个对于肌肉感应钙离子非常重要的蛋白质。一个肌钙蛋白分子与所有的原肌球蛋白分子作用。肌钙蛋白是一个三条链的复合体，即肌钙蛋白 C、肌钙蛋白 I 和肌钙蛋白 T。肌钙蛋白 C 是一个有 4 个 EF 手结构的钙结合蛋白，它与其他蛋白质结合形成一个用于调控的头部。其他的两个分子形成长的螺旋（图 12.13）。这个复合体调控着位于肌动蛋白纤维上的肌球蛋白结合位点，控制其在肌肉收缩过程中是否可以被肌球蛋白所接触到。

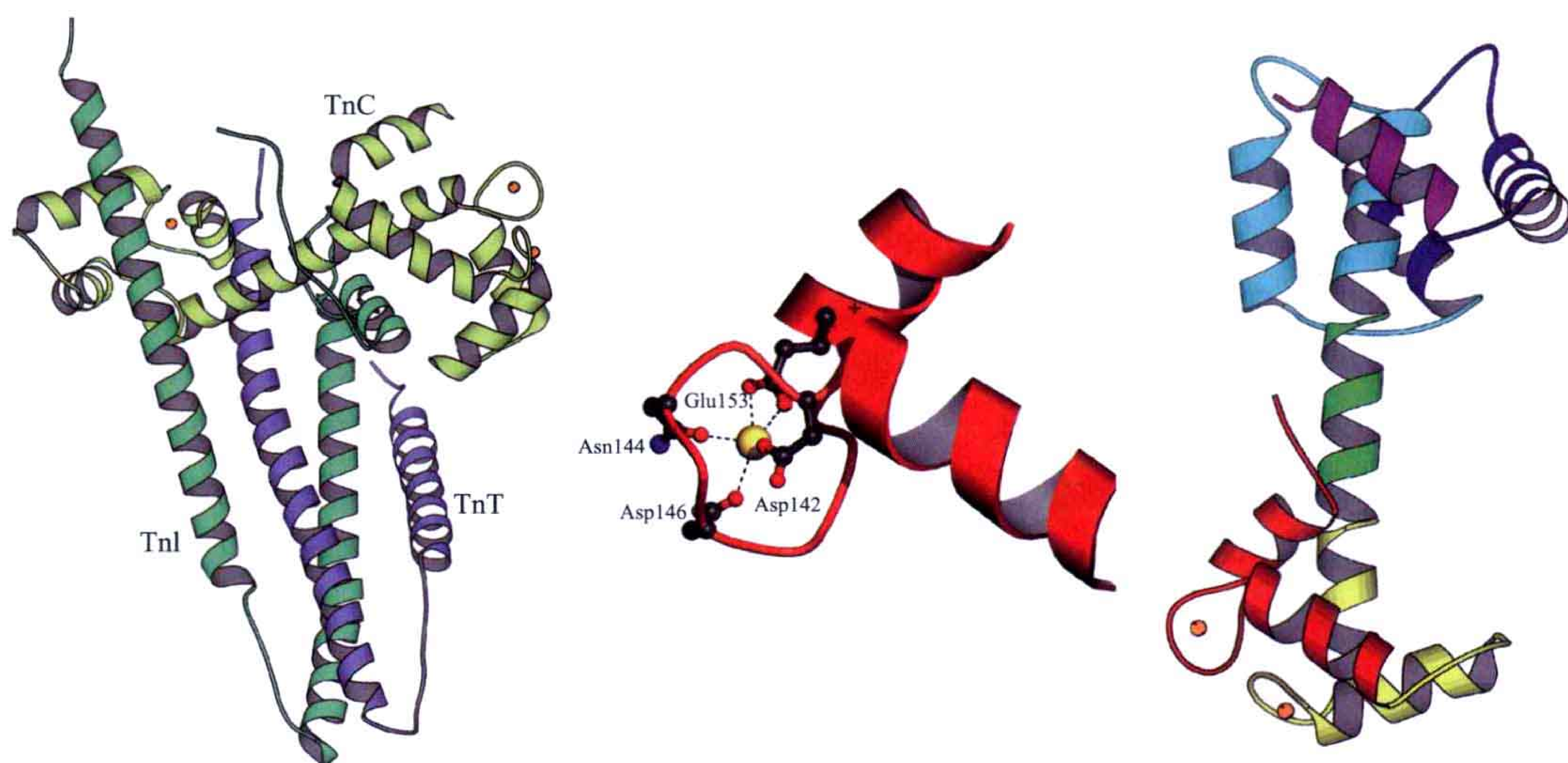


图 12.13　左：肌钙蛋白复合体。肌钙蛋白 I 和肌钙蛋白 T 的 C 端形成像筷子一样的双螺旋（PDB：1YTZ）。肌钙蛋白 C 的位置则像是操作筷子的手一样。右：肌钙蛋白 C 是一个典型的 EF 手蛋白，两对 EF 手之间由一个长的螺旋结构相连。EF 手的颜色分别为蓝色、青色、黄色和红色。在这个晶体结构中，只有 C 端的一对 EF 手结合了钙离子（PDB：5TNC）。中：其中一个 EF 手结构与钙离子结合的细节图（即右图中的红色部分）。来自天冬氨酸、天冬酰胺和谷氨酸侧链的氧原子都与钙离子有相互作用，同样与钙离子相互作用的还有一个羰基氧及一个水分子（图中示显示）。

12.2.3　肌球蛋白

肌球蛋白是一种 ATP 水解酶，可以移动肌丝，它是一种机械化学酶及分子马达蛋白。它们利用 ATP 的化学能做功来完成分子水平的运动，最终导致宏观的运动。

人们研究得最多的肌球蛋白称为肌球蛋白 Ⅱ，它存在于动物之中，同样也存在于

简单的生物体如酵母之中。这是一个在肌丝的粗丝中发现的分子，但是它也在其他类型的细胞中存在。肌球蛋白Ⅱ由6条链组成：两条约1950个氨基酸的重链及两对较小的蛋白质。

在人类基因组中有超过40种不同的肌球蛋白重链基因（表12.2）。另外还有一些肌球蛋白Ⅱ、肌球蛋白Ⅰ和肌球蛋白Ⅴ的基因，肌球蛋白Ⅰ和肌球蛋白Ⅴ与细胞膜相关，是细胞活性骨架的一部分。肌球蛋白Ⅱ和Ⅴ中有两条重链，而肌球蛋白Ⅰ则以一条重链行使功能。

表 12.2　一些类型的肌球蛋白及其功能

类型	重链的分子质量 /kDa	主要功能
Ⅰ	110 ～ 150	结合细胞膜
Ⅱ	220	肌肉中的肌丝滑动
Ⅴ	170 ～ 220	囊泡运输
Ⅵ	140	运输胞吞的囊泡，沿肌动蛋白纤维的“－”末端方向移动

所有肌球蛋白的重链都与EF手型的钙结合蛋白相互作用。在肌球蛋白Ⅱ中，重链与调控轻链（regulatory light-chain, RLC）和基本轻链（essential light-chain, ELC）作用；而在肌球蛋白Ⅰ、Ⅴ和Ⅵ中，重链与钙调蛋白作用。

每一个肌球蛋白都有一个头部（或者称为马达结构域）、一个颈部及一个C端的尾部（图12.14）。尾部代表的是一个蛋白质的二聚体，在肌球蛋白Ⅰ和Ⅴ中，它控制着肌球蛋白与膜蛋白的相互作用。尾部的大部分含有一个28个氨基酸的重复序列。这与预想中的尾巴构象是一致的：两条重链形成一个由两个长α螺旋相互缠绕而成的卷曲螺旋。肌球蛋白Ⅴ和Ⅵ在C端有一个可以用来结合货物（囊泡或者细胞器）的结构域。

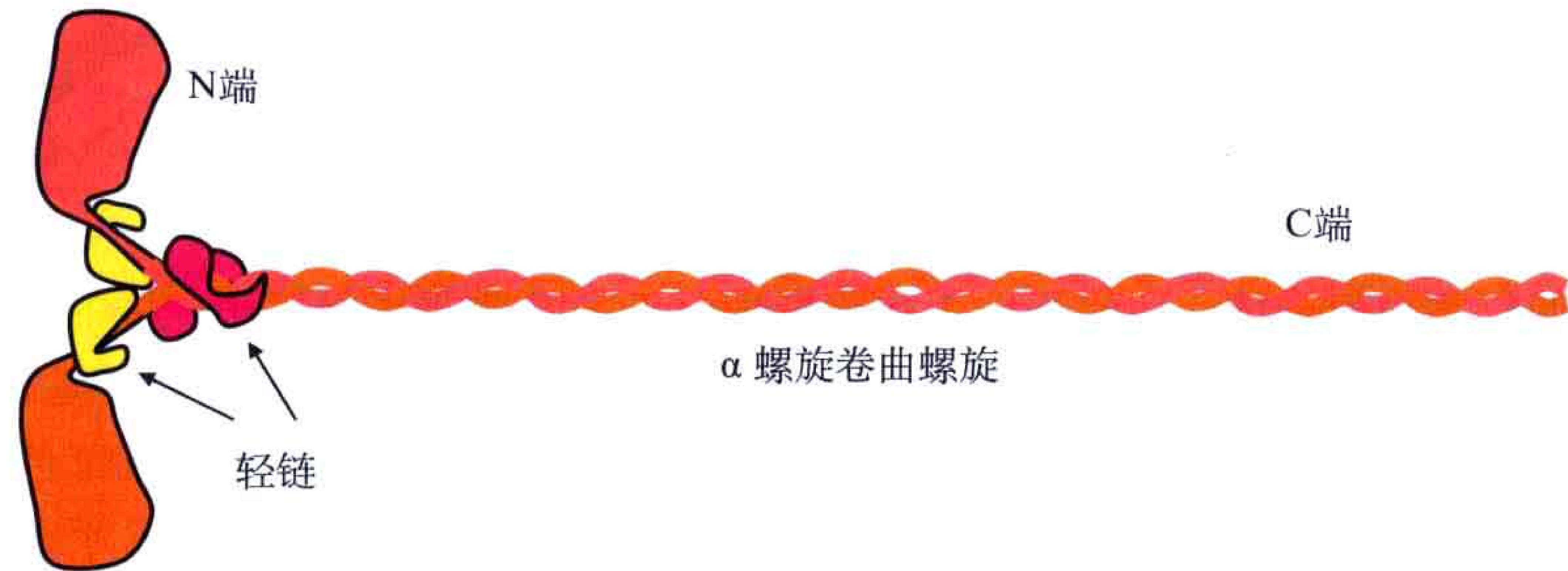


图 12.14　肌球蛋白Ⅱ分子的示意图。它有两条重链及两对轻链。

Ⅰ、Ⅴ和Ⅵ型肌球蛋白与肌球蛋白Ⅱ的马达结构域十分相似，序列差异主要出现在尾部。

12.2.3.1　肌球蛋白Ⅱ

目前被解析的肌球蛋白Ⅱ的晶体结构是它的一个被称为S1的片段（图12.15）。这个片段包含了头部与颈部的区域：C端尾部重链上的约1000个氨基酸的区域被蛋白

酶切除了。颈部包含了两条轻链和被结晶的重链的 C 端。在这个晶体结构中重链形成了一个非常长且弯曲的螺旋，轻链则围绕着这个螺旋排列。两条轻链的组合方式均与钙调蛋白和肌钙蛋白 C 相同，即有两个球状结构域，每个含两个 EF 手，两个结构域之间由一段柔性的、部分螺旋的片段连接。调控轻链和基本轻链上的 EF 手中仅有少数可以与钙离子结合。

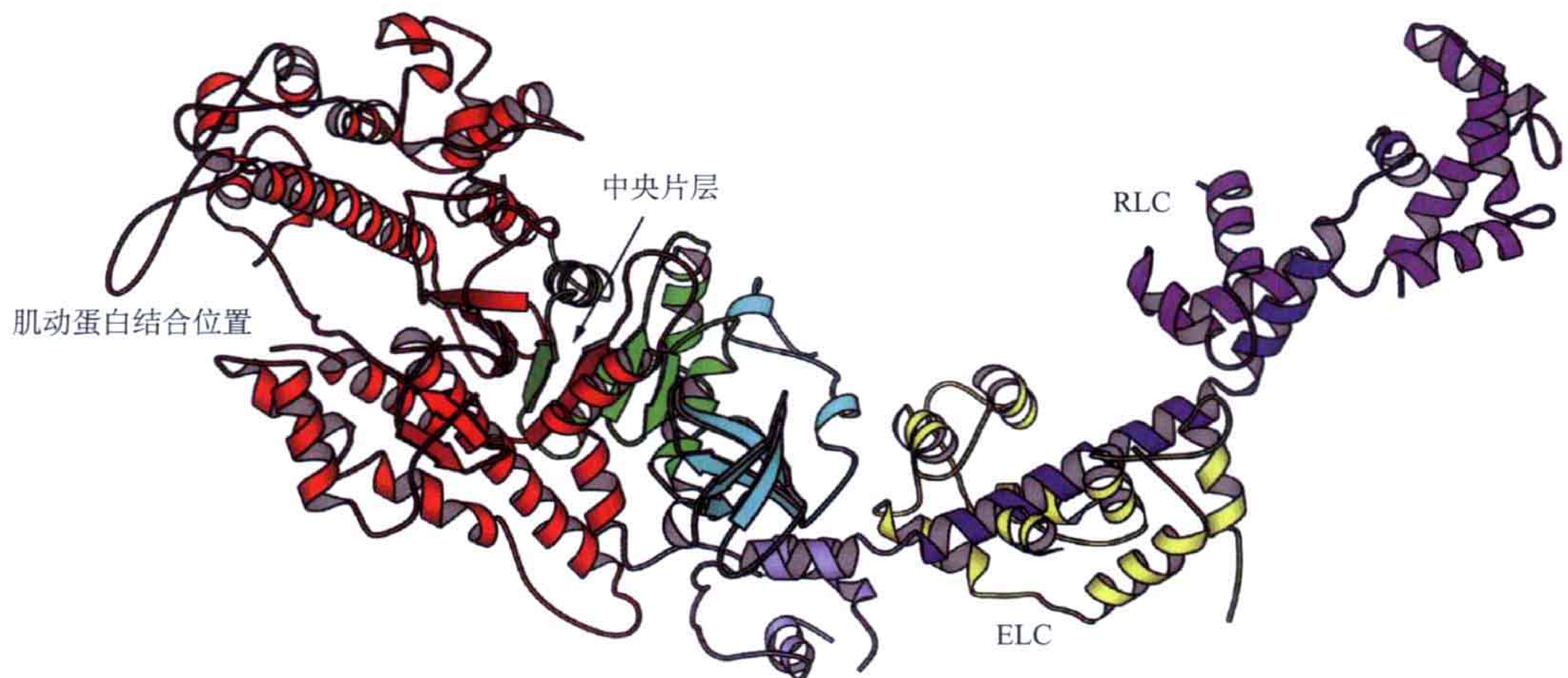


图 12.15　鸡肌球蛋白 S1 片段（头部及颈部）的结构。其中包含了重链的 N 端及两条轻链（PDB：2MYS）。头部有 5 个结构域：N 端的桶状结构域（青色）、中央结构域（绿色）、上下 50K 结构域（红色）、转化结构域（亮蓝色）及 C 端的螺旋（深蓝色）。ELC（黄色）即基本轻链，RLC（紫色）是调控轻链。两条轻链都是由两个球状结构和一个连接段组成，每一个球状结构域上有两个 EF 手。ATP 分子结合在中央片层上。

肌球蛋白的头部是一个非常复杂的结构，它可以被分为几个结构域。在它的 N 端，有一个小的反平行 β 片层，或者称为 β 桶（即 N 端结构域）。而头部的中央部分也是一个 β 片层，是由来自蛋白质链上三个不同片段的折叠束组成的。这个片层主要是平行排布的。连接折叠束的主要是螺旋，其中一些螺旋形成了一个独立的结构域，位于整个分子的顶端。它们被称为上 50K 和下 50K 结构域，同时头部剩下的区域也被称为下 50K 结构域。

12.2.3.2　肌球蛋白结构与 G 蛋白的关系

中央片层的中央折叠束以一个在许多核苷酸结合蛋白中都被发现的序列结尾。这个序列模式是 GXXXXGKS/T，是这些蛋白质 P-loop 的一部分。ATP 分子便结合在这个位置，与 GTP 结合 G 蛋白的 P-loop 的方式很相似（图 12.16）。

肌球蛋白中央片层的拓扑结构与 G 蛋白的并不相同，但也可能是因为肌球蛋白与 G 蛋白有比较远的进化上的关系。两者除了 P-loop 的相似度之外，位于中央折叠束两边且位于 P-loop 之前的两个折叠束（肌球蛋白的折叠束 6 和 7 与 Ras 的折叠束 3 和 4）在序列上也位于同样的位置。片层中的其他折叠束则来自不同的片段。

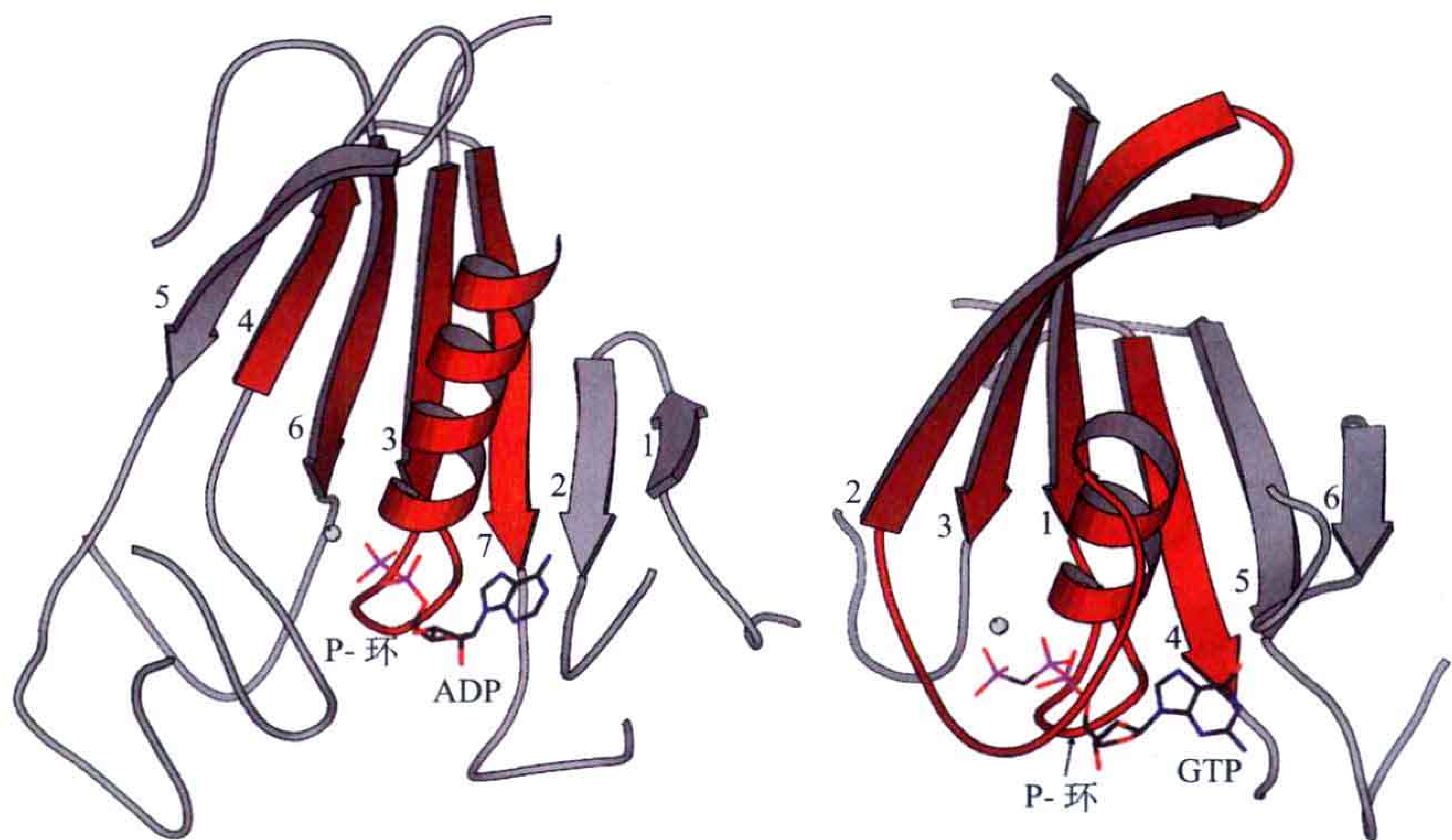


图 12.16 肌球蛋白（黏液菌中的，PDB：1MMA，左）与 G 蛋白 Ras（右）中折叠片层和 P-loop 的比较。图中数字表示的是 β 折叠束的顺序。两者只有一部分折叠束是相同顺序的（红色）。肌球蛋白中的上、下 50K 结构域是一个很大的插入结构域，插入了折叠束 5 和 6 及 6 和 7 之间。

12.2.4 肌肉收缩的机制

当肌肉收缩时，肌球蛋白使肌丝做相对移动。这个工作的能量来源于肌球蛋白中 ATP 的水解。如何将 ATP 的化学能转化为动能的细节目前尚不清楚。粗丝是具有两极的，肌球蛋白的头部在粗丝的两个末端排成螺旋状。粗丝的头部通过结合肌动蛋白纤维与细丝形成交联桥。ATP 的水解导致了肌球蛋白颈部的一个较大的构象变化。肌球蛋白的尾部则像杠杆臂一样活动（“摆动的杠杆臂模型”）并导致肌丝相对滑动，将粗丝移向肌动蛋白纤维的“+”末端从而缩短肌小节。肌球蛋白的头部随后从肌动蛋白纤维上释放下来从而可以结合在新的位置上以重复这个循环。因为肌丝上有大量的肌球蛋白头部，有一些结合有一些分离，于是打滑的风险是很低的。

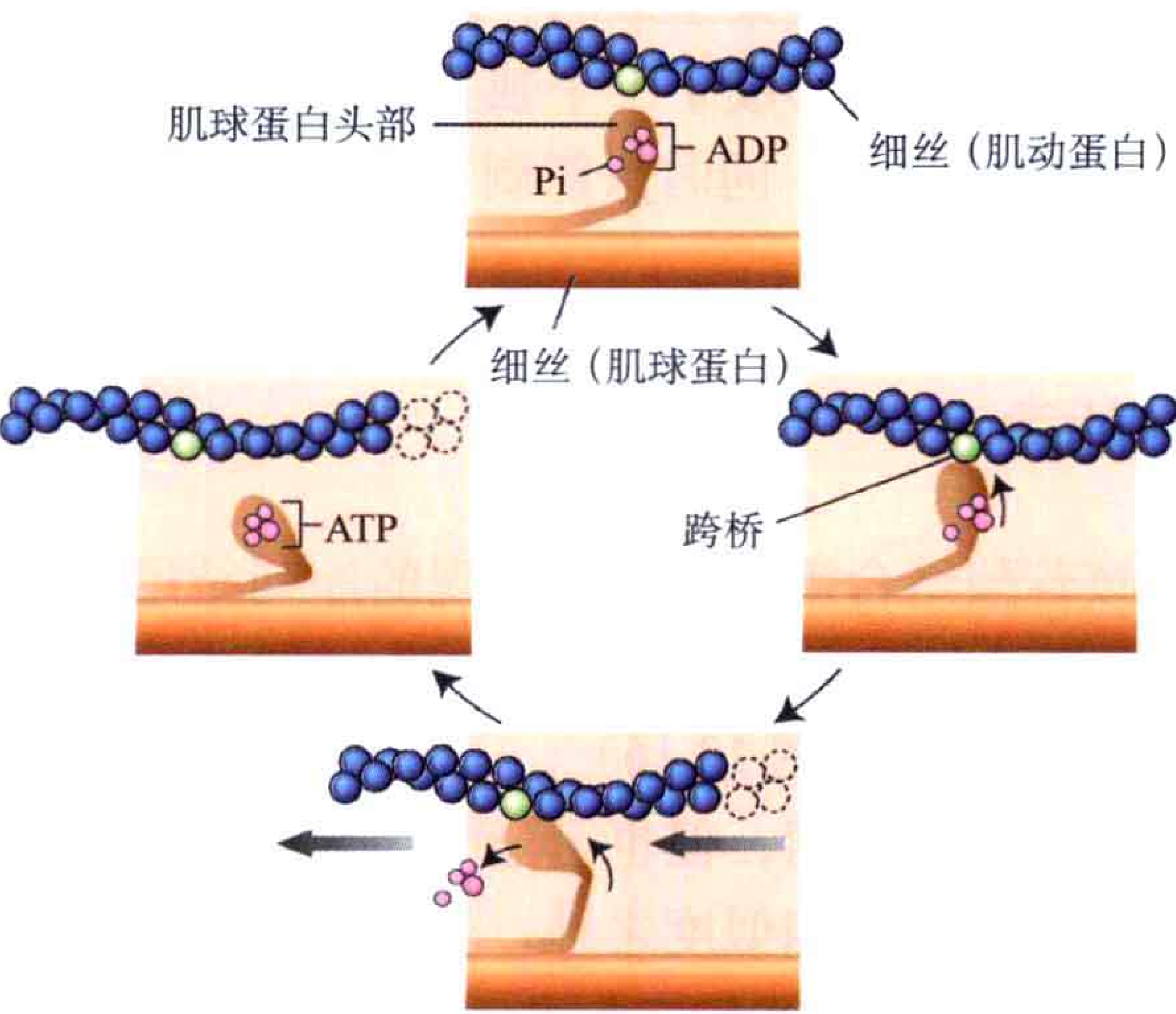


图 12.17 肌肉收缩步骤图示。

当 ATP 不在时，肌球蛋白非常紧地结合在肌动蛋白纤维上（“僵硬”状态）。根据假说，肌球蛋白的循环涉及下列步骤（图 12.17）。

(1) ATP 结合导致了构象变化，从而使肌球蛋白与肌动蛋白之间的相互作用被破坏。

(2) ATP 水解导致构象变化，然后肌球蛋白结合在肌动蛋白纤维的新

位置上。注意 ATP 的水解有可能发生在肌球蛋白结合肌动蛋白纤维之前。

(3) 磷酸从肌球蛋白上离去，导致另一个构象变化。

(4) 最终，ADP 从肌球蛋白上解离，于是肌球蛋白恢复到休止状态或者说是“僵硬”状态，准备从步骤（1）重新开始。纤维的真正运动（“做功冲程”）发生在步骤（3）或（4）。

通过将肌球蛋白头部及肌动蛋白单体的原子结构模型放入被肌球蛋白头部“修饰”后的肌动蛋白纤维的电子显微镜密度中，可以很好地确定出肌球蛋白头部的肌动蛋白结合裂隙（图 12.18），但电镜的分辨率无法支持对其相互作用表面更细节的研究。



图 12.18　肌动蛋白单体与鸡肌球蛋白 S1 的结构和“修饰”后的肌动蛋白电镜密度的匹配图。图中显示了三个肌动蛋白单体。马达结构域和一个单独的轻链（图右）只独立的刚性体进行拟合匹配 [此图原出自 Volkmann N, Hanein D. (2000) Actomyosin: Law and order in motility. *Curr Op Cell Biol* **12:** 26-34. Elsevier 版权所有，感谢 Hanein 和 Volkmann 博士授权刊印]。

许多肌球蛋白片段与 ADP 或 ATP 类似物结合或非结合的结构已被解析，这些结构显示出肌球蛋白的不同构象。但是，并不能由此推论观察到的构象可等同于肌球蛋白循环中的构象，况且一些预期的状态还没有得到结构。晶体中的肌球蛋白头部的构象并不对结合核酸敏感。一个可能的解释是我们观察到的构象与肌肉组织中的并不相同，肌球蛋白与肌动蛋白的相互作用可能对于循环中的某种构象变化的发生是至关重要的。我们还可以看出肌球蛋白是一个拥有非常复杂的构象调控机制的蛋白质。肌球蛋白与肌动蛋白在头部顶端的相互作用及与 ATP/ADP 在头部中部的相互作用可以使颈部和远

离头部的末端产生构象变化。

虽然肌球蛋白的结构显示其头部大部分区域仅有较小的构象变化，但其颈部区域及相连的卷曲螺旋区域还是显示出非常不同的构象（图 12.19）。头部区域的各结构域在不同的结构中只有较小的移动。发生构象变化的部分有：连接上、下 50K 结构域的开关 II loop、一个位于两个长螺旋之间的连接 loop 及 SH1 螺旋，其中 SH1 螺旋的作用是将下 50K 结构域连接到转换结构域。转换结构域是一个放大器，位于 C 端长螺旋的起始部分（图 12.20），它将头部的构象变化传递给颈部和尾部。肌球蛋白的卷曲螺旋起一个杠杆臂的作用，用于放大发生在头部的微小的构象变化。

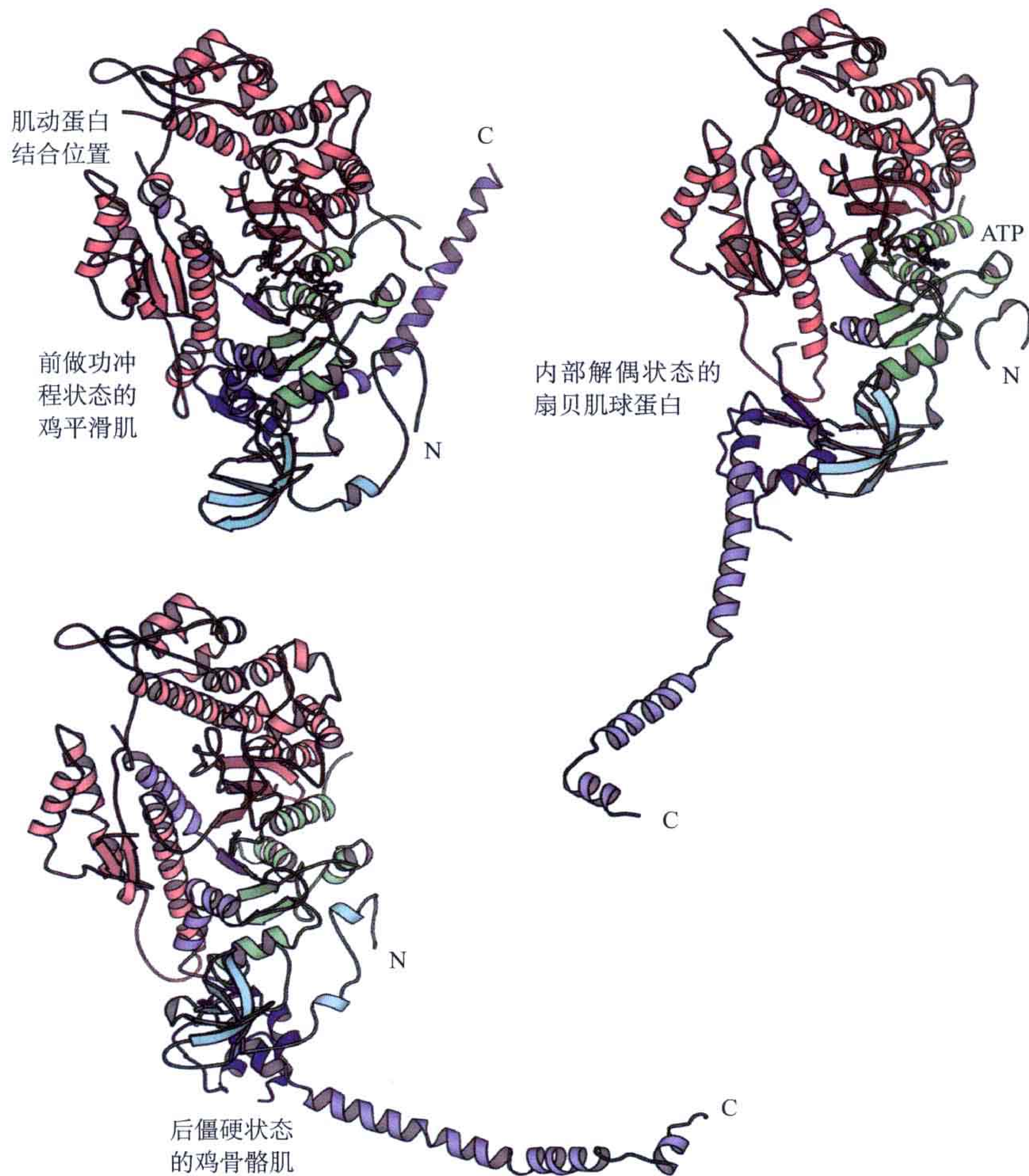


图 12.19　肌球蛋白 S1 片段被观察到的三种主要构象。上左：鸡平滑肌（chicken smooth muscle）肌球蛋白（结合 ADP 和 AlF4，PDB：1BR1），前做功冲程（pre-power stroke）状态；上右：扇贝（scallop）肌球蛋白（结合 ADP，PDB：1B7T），内部解偶（internally uncoupled）状态；下图：鸡骨骼肌（chicken skeletal muscle）肌球蛋白（无核苷酸结合，PDB：2MYS），后僵硬（post-rigor）状态。三种构象的头部都处于相同的朝向，颈部则可以看到很明显的不同朝向。后僵硬状态大致上等同于循环结束时杠杆臂的朝向，而前做功冲程状态则接近于头部重新与肌动蛋白纤维结合前的状态。

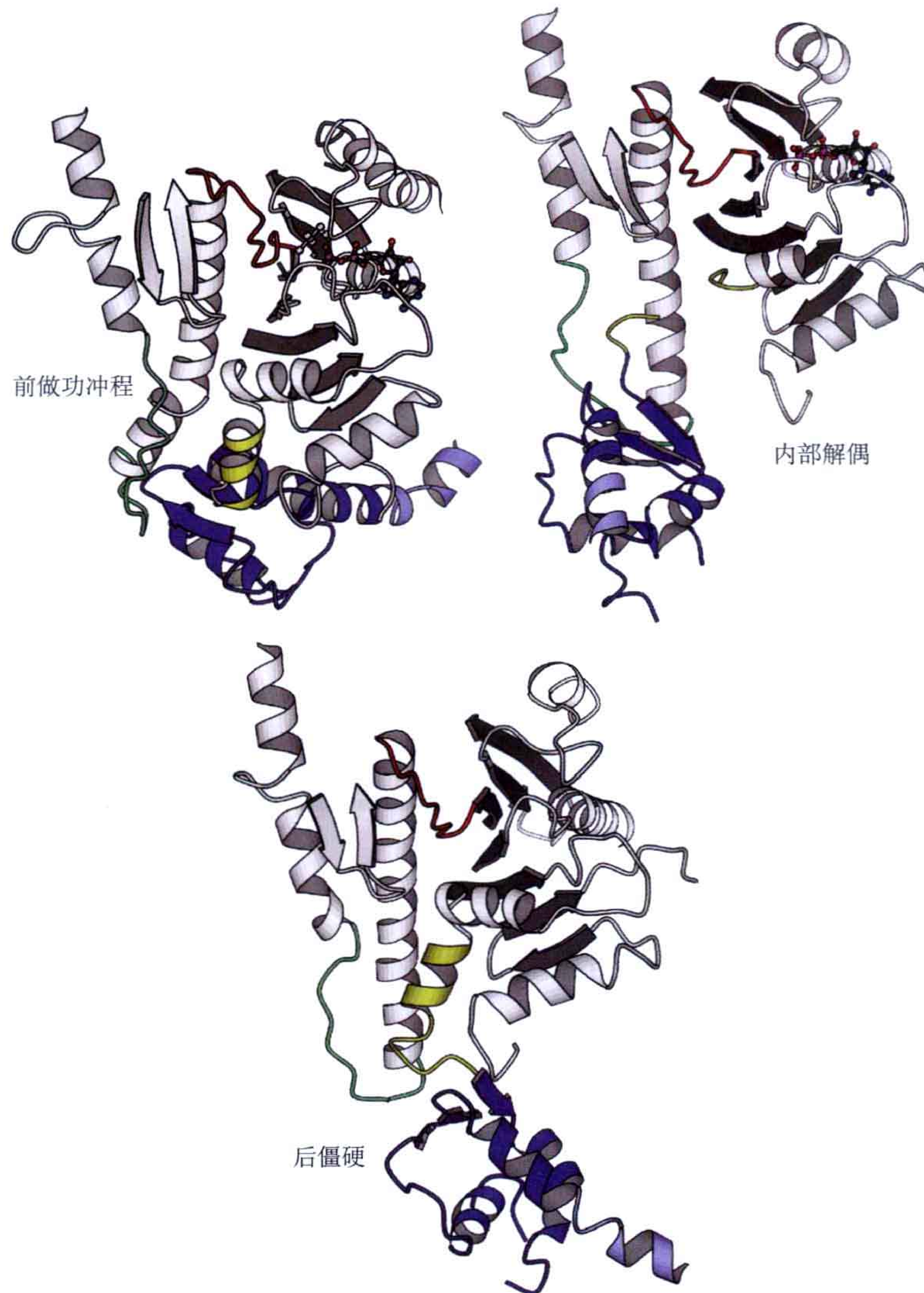


图 12.20　肌球蛋白头部的构象变化引发尾部朝向变化的细节图。中央片层为灰色。下：鸡骨骼肌肌球蛋白；上左：鸡平滑肌肌球蛋白；上右：扇贝肌球蛋白。连接亚结构域的片段有一些构象变化：开关Ⅱ（红色）、SH1 螺旋（黄色，在扇贝肌球蛋白的结构中是无序的）及连接 loop（绿色）。开关Ⅱ的构象不同，伴随着垂直的长螺旋以及转化结构域的变化。这些片段对于与 ADP 和 ATP 的结合，以及和肌动蛋白纤维的结合是敏感的。后僵硬状态和前做功冲程状态比较接近于构象变化开始和结束的边界构象。

至今一共观察到三种主要的肌球蛋白的构象，分别是鸡骨骼肌肌球蛋白、鸡平滑肌肌球蛋及扇贝肌球蛋白。这三个结构被认为代表了三种状态，分别称为后僵硬状态、前做功冲程状态或称中间状态、内部解偶状态或称分离状态。后僵硬状态和前做功冲程状态比较接近于构象变化开始和结束时的边界构象。在前做功冲程状态时，杠杆臂处于预期中肌球蛋白头部结合在肌动蛋白纤维上且 ATP 已被水解的位置。在后僵硬状态时，杠杆臂处于运动（做功冲程）已经发生后的位置。

有可能这些结构都无法等同于一个对肌动蛋白有高亲合力的构象。位于上、下 50K

结构域之间的肌动蛋白结合裂隙在肌球蛋白结合肌动蛋白时也许要闭合得更紧一些。核苷酸结合位点与肌动蛋白结合位点在结构上的联系可能与 G 蛋白（如 Ras）类似。肌球蛋白上对应于 G 蛋白中开关区的区域在存在或不存在 γ 磷酸时发生了构象变化，而这个构象变化也许影响了其与肌动蛋白的结合。上 50K 结构域的移动偶联着结合核苷酸的 loop 的移动与构象变化。而上、下 50K 结构域的靠近则增加了肌球蛋白对肌动蛋白的亲合性并导致核苷酸结合口袋的打开及核苷酸的释放。肌球蛋白结合 ATP 导致 50K 结构域的间隙打开从而削弱其与肌动蛋白的结合。

12.2.4.1 肌动蛋白与肌球蛋白在细胞运输中的作用

肌动蛋白纤维同样参与细胞运输及细胞极性的发展。微管参与长距离的运输；肌动蛋白与肌球蛋白（Ⅵ类）则参与运输细胞物质，如线粒体、mRNA 到合适的附近位置。病毒的组分也可以劫持肌动蛋白的运输系统以便快速地聚集新的病毒体。

12.3 微管

12.3.1 微管蛋白的结构与功能

微管是一种管状的结构，在细胞中有两个功能：一个功能是胞质中囊泡的细胞内运输；另一个功能是参与细胞分裂，即负责分开已复制的染色体。微管主要由一个蛋白质即微管蛋白组成。微管蛋白有两类亚基，即 α 亚基和 β 亚基，它们形成异源二聚体。这些二聚体以头对尾的方式排列成原丝（protofilament）。13 根原丝侧面与侧面规则地相互作用并形成微管（图 12.21）。这个管状结构比肌动蛋白纤维更硬，被用于行使一些更需要机械强度的功能。微管的生长和解离方式与肌动蛋白纤维类似，这些过程被其他的蛋白质所调控。

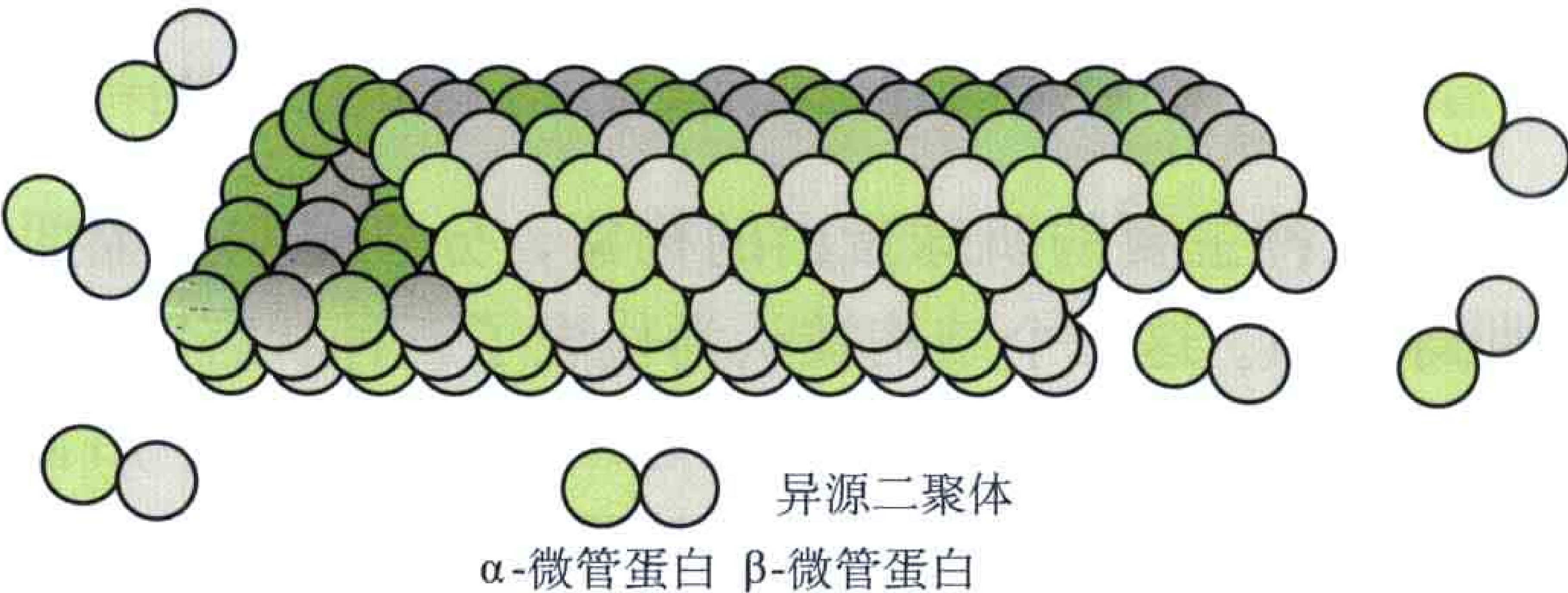


图 12.21　微管示意图。

α 亚基和 β 亚基单体都结合 GTP，β 亚基有 GTP 水解酶的活性。在微管生长端，

所有的单体都含有 GTP，但其他部分 β 亚基中的 GTP 已被降解成了 GDP。微管末端的 GTP 含量决定了它处于生长状态还是解聚状态。

微管蛋白的结构已通过电子晶体学（electron crystallography）解析。微管蛋白可以被诱导形成片层，这些片层便是二维的晶体。在电子显微镜下，这些晶体可以用来生成分辨率相对较高的电子密度图。微管蛋白的片层由原丝组成，原丝之中的相互作用因此也被保留。然而，原丝在晶体中是以反平行的方式排列，而在生物体内它们是平行排列的，所以在晶体结构中观察到的原丝间相互作用与生物体内的结构并无关联。

α 亚基和 β 亚基在构象上十分相似（图 12.22）。它们都是一个由三个结构域组成的紧密单元。第一个结构域是 GTP 结合结构域，它有经典的 Rossmann 折叠，即由螺旋连接的 6 个平行的折叠束，这与许多脱氢酶都类似。第二个结构域是由平行或反平行的折叠片层及两侧的螺旋形成的。第三个结构域是两个螺旋，它们和第一个结构域相对排列。GTP 结合结构域的拓扑结构与 G 蛋白的相似，但不完全相同，在第一个 β 折叠束之后的 loop，即对应于 G 蛋白中 P-loop 的 loop 上并没有 GXXXXGKT/S 的序列。核苷酸的结合是靠这个 loop 与另外一个富含甘氨酸的 loop 维系的，这个 loop 在微管蛋白分子中是保守的。

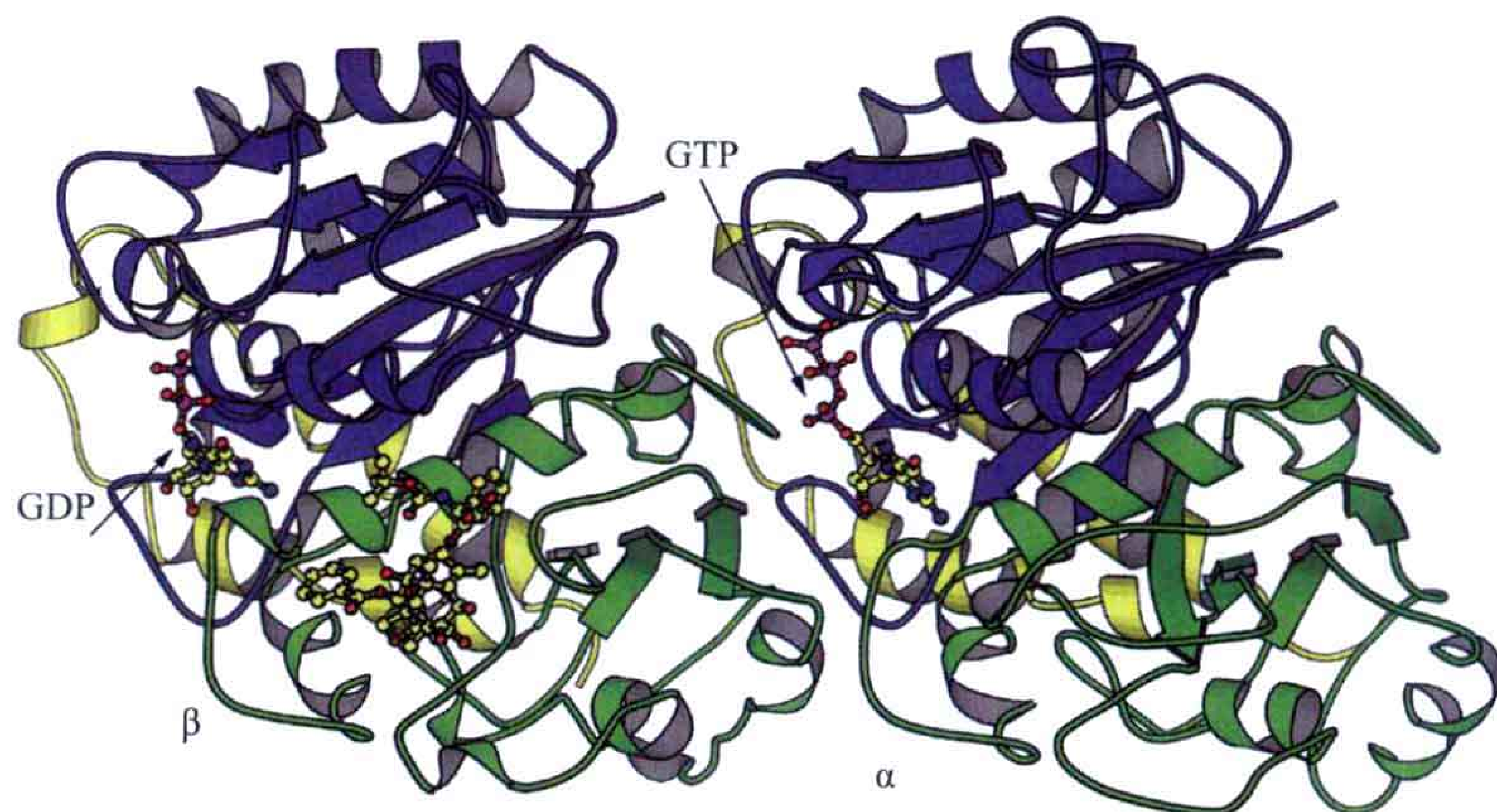


图 12.22　微管蛋白 αβ 二聚体的结构。Rossmann 折叠结构域（蓝色）位于上方。第二结构域显示为绿色，C 端的螺旋显示为黄色。解析的结构中包含了一个结合在 β 蛋白上的药物分子（PDB: 1TUB）。

从结构中可以看出，GTP 水解在 α 和 β 亚基中的不同。在二聚体中，α 亚基的 GTP 位点被封闭在 α 和 β 亚基的界面中；β 亚基的位点则暴露在水环境中，使得水解反应可以被 α 亚基所激发。

微管生长的起始依赖第三种微管蛋白，即 γ 微管蛋白，它与另两种微管蛋白亚基有着相似的结构与 GTP 水解酶活性。有一些分子可以稳定微管蛋白纤维，如微管结合蛋白（microtubule-associated protein，MAP）。另一些分子则具有去稳定性的功能，如药物秋水仙碱（colchine）及胞质磷蛋白（stathmin）家族的一些蛋白质。

微管蛋白与秋水仙碱及类胞质磷蛋白的复合体的结构显示了秋水仙碱与胞质磷蛋

白如何去稳定化微管蛋白（图 12.23）。RB3 蛋白的类胞质磷蛋白结构域有一个 C 端的长螺旋，它与两个微管蛋白二聚体结合，这两个微管蛋白二聚体以头到尾的方式排列，且略有弯曲。微管蛋白原丝的这个弯曲的形式使其不适合形成微管，这种结合也会促使微管蛋白的解聚。秋水仙碱与 β 亚基结合，靠近微管蛋白二聚体的界面。

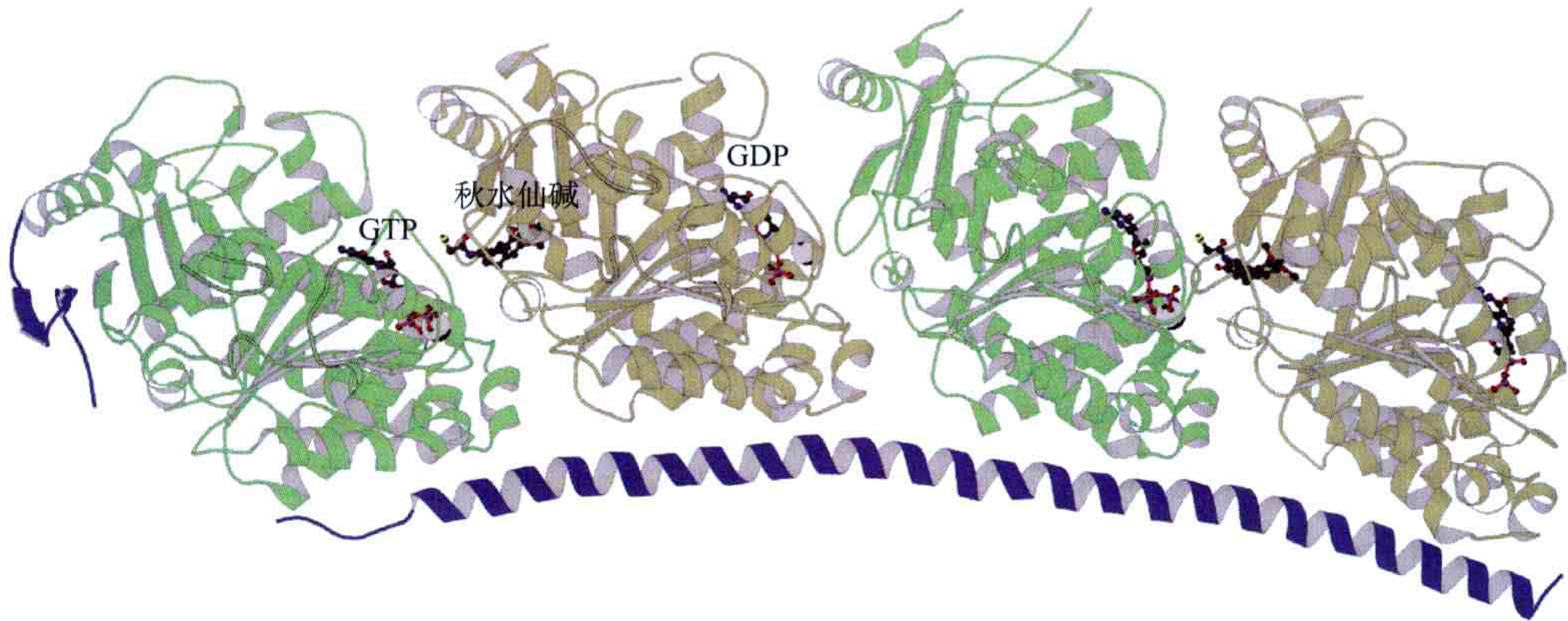


图 12.23　两个微管蛋白异二聚体、RB3 的类胞质磷蛋白结构域（蓝色）和秋水仙碱的复合体。α 亚基显示为绿白色，β 亚基显示为棕白色。胞质磷蛋白结构域形成一个长的螺旋及一个小的 β 发夹结构，但这两个结构之间的结构是无序的。N 端的发夹结构封闭了微管蛋白的一个末端；C 端的长螺旋沿着略弯曲的微管蛋白亚基排列（PDB：1SA0）。

12.3.2　与微管相关的马达蛋白：驱动蛋白和 Ncd

囊泡沿微管的运输是利用两种能产生动力的蛋白质——动力蛋白（dynein）和驱动蛋白（kinesin）完成的。驱动蛋白是一种与肌球蛋白相似的马达蛋白。驱动蛋白家族可分为三种类型，即 N 型、C 型和 M 型，这取决于其马达结构域的位置（N 端、C 端和中间）。N 型驱动蛋白被用于沿着微管进行运输。它们利用 ATP 水解的能量来产生动力，以移动细胞器或染色体。绝大多数驱动蛋白都是二聚体，有两个相同的重链，每个重链约 1000 个氨基酸。像肌球蛋白一样，N 型驱动蛋白由一个 N 端的马达结构域（头部）和一个长的螺旋区域形成二聚体（杆部）。被运输的货物是被黏附在轻链上，这主要是通过货物与尾部螺旋 C 端相互作用实现的。

12.3.2.1　驱动蛋白的结构和功能

晶体结构显示 N 型驱动蛋白 N 端的马达结构域与肌球蛋白的头部很相似，但要小很多（请比较图 12.24 与图 12.16）。它由一个主要是平行排列的片层和围绕在两侧的螺旋构成。N 型驱动蛋白之所以与肌球蛋白头部的大小有差异主要是因为肌球蛋白在一些位置有大的插入结构域，这些插入结构域形成了与肌动蛋白结合的裂隙。

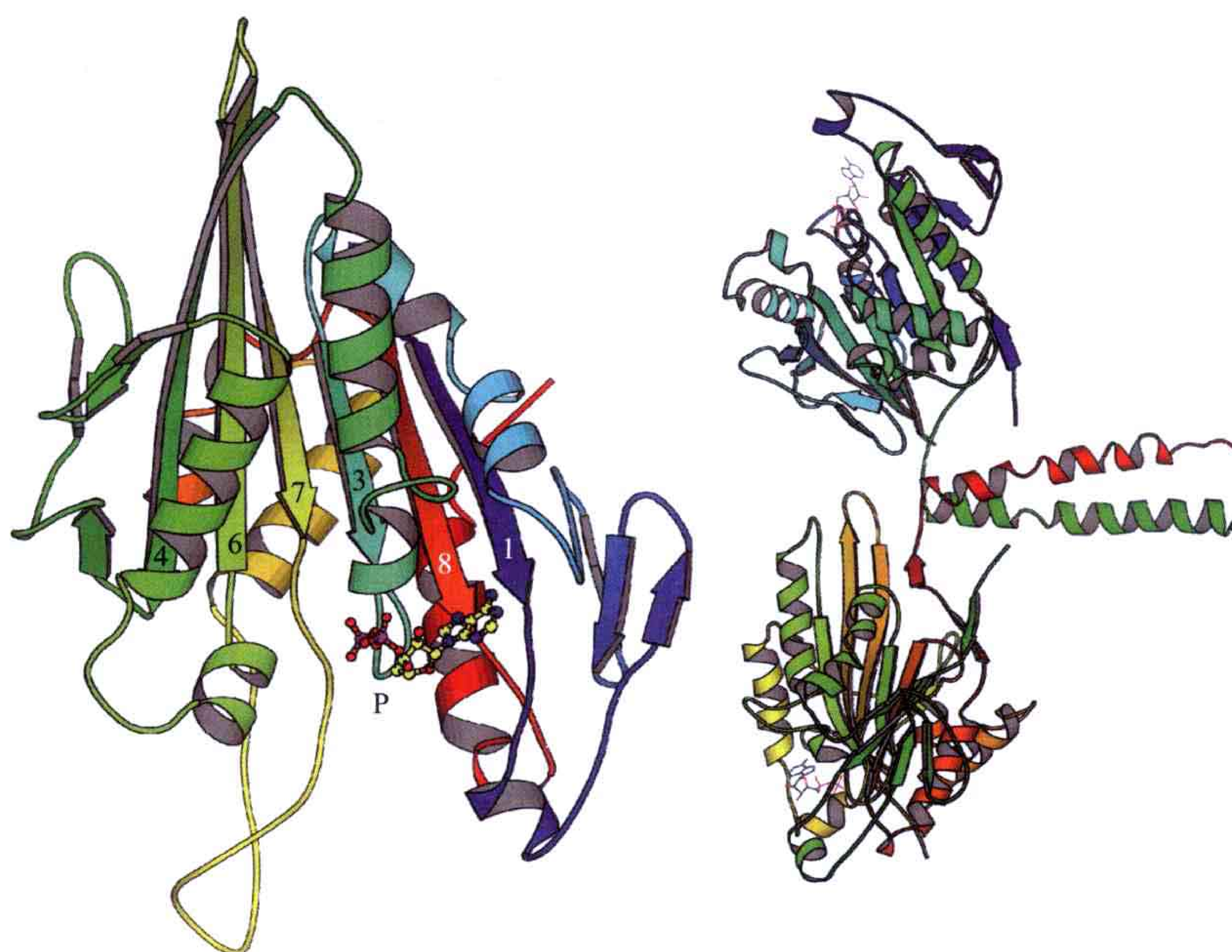


图 12.24　左：驱动蛋白马达结构域的结构（N 型）。P-loop 在磷酸基团的附近标示（PDB：1BG2）。右：驱动蛋白二聚体及其螺旋状的颈部区域（PDB：3KIN）。在螺旋颈部的末端还有约 600 个氨基酸的杆部。

从微管与驱动蛋白复合体的电镜信息及驱动蛋白的晶体信息，我们可以发现驱动蛋白的两个头部是同时结合在微管上的。在移动时，一个头部绕着另一个头部旋转从而结合到微管上新的位置。用这种方式，驱动蛋白二聚体可以驱动一个连接在其杆部的物体，如细胞器的运动。目前同样不清楚驱动蛋白产生动力的机制，但是其头部结构域的构象变化应该跟肌球蛋白的类似。肌球蛋白中被用来移动杠杆臂的转化结构域在驱动蛋白结构中并没有明显对应的结构域，因此并不清楚预想中较小的构象变化如何可以转换为大幅度的运动。驱动蛋白中，连接卷曲螺旋和马达结构域的颈部是一个 15 个氨基酸的保守片段，它可能行使着转换功能。

另一个与微管相关的马达蛋白被称为 ncd（non-claret disjunctional），它在染色体分离时很活跃。这个蛋白质属于 C 型的驱动蛋白，它的马达结构域位于蛋白质的 C 端。它与驱动蛋白十分相似，但以一个完全相反的方向行动，即朝向微管的“–”末端移动。这个差异很可能与马达蛋白在蛋白质链中的位置有关。第三种类型的驱动蛋白，M 型或者称为 KinI 驱动蛋白，有着相似的马达结构域，但这类蛋白质并不沿着微管运输物质，它们的功能是结合并且解聚微管。其催化核心的结构显示这类蛋白质与其他驱动蛋白是相似的。对于其他的驱动蛋白来说，由于缺少与微管的复合体结构，所以了解它们的工作机理非常困难。

延伸阅读

原始文献

Holmes KC, Popp D, Gebhard W, Kabsch W. (1990) Atomic model of the actin filament. *Nature* **347**: 44-49.

Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC. (1990) Atomic structure of the actin: DNase 1 complex. *Nature* **347**: 37-44.

Nogales E, Wolf SG, Downing KH. (1998) Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**: 199-203.

Rayment I, Rypniewski WR, Schmidt-Bäse K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM. (1993) Three-dimensional structure of myosin subfragment-1: A molecular motor. *Science* **261**: 50-58.

Robinson RC, Mejillano M, Le VP, Burtnick LD, Yin HL, Choe S. (1999) Domain movement in gelsolin: A calcium-activated switch. *Science* **286**: 1939-1942.

Robinson RC, Turbedsky K, Kaiser DA, Marchand JB, Higgs HN, Choe S, Pollard TD. (2001) Crystal structure of Arp2/3 complex. *Science* **294**: 1679-1684.

综述文章

Amos LA, Schlieper D. (2005) Microtubules and MAPs. *Adv Prot Chem* **71**: 258-298.

Craig R, Woodhead JL. (2006) Structure and function of myosin filaments. *Curr Opin Struct Biol* **16**: 204-212.

Geeves MA, Holmes KC. (2005) The molecular mechanism of muscle contraction. *Adv Prot Chem* **71**: 161-193.

Marx A, Muller J, Mandelkow EM, Hoenger A, Mandelkow E. (2006) Interaction of kinesin motors, microtubules, and MAPs. *J Muscle Res Cell Motil* **27**: 125-137.

Winder SJ. (2003) Structural insights into actin-binding, branching and bundling proteins. *Curr Opin Cell Biol* **15**: 14-22.

（曹 駸 译 黄亦和 苏晓东 校）

第十三章
细胞间相互作用的结构基础

13.1 细胞外基质蛋白

在动物组织中，细胞被有序地组织在称为细胞外基质（extracellular matrix，ECM）的网络中。细胞外基质的主要成分是胶原蛋白纤维、蛋白聚糖和各种可以与其他成分结合的胞外基质蛋白。不同的组织由不同的基质构成，使其具有相应的特征。细胞外基质的成分是由组织内的细胞合成的，并处在持续更新中。

13.1.1 胶原蛋白纤维

胶原蛋白是纤维状蛋白，是很多组织的主要成分，它们主要起结构作用，很多蛋白质可以与组织和细胞外基质中的胶原蛋白结合。

编码胶原蛋白多肽的基因有很多，生成的不同胶原蛋白多肽可以组合形成不同类型的胶原蛋白，以满足不同组织的需要。在细胞中，胶原蛋白以前体分子的形式，大多数类型胶原蛋白纤维是通过去除前体蛋白的 N 端或 C 端形成的。每条这种胶原蛋白链有 1050 个氨基酸残基，三个胶原蛋白分子相互缠绕成为三股螺旋，形成原胶原（tropocollagen）；原胶原错位交叉使纤维加长，形成原纤维。胶原蛋白三螺旋通过链末端的赖氨酸侧链共价连接，这些赖氨酸被细胞外基质的赖氨酸氧化酶氧化成醛，醛基与相邻链的赖氨酸发生反应形成交联。

胶原蛋白是 X-Y-Gly 序列的高度重复，其中 X 和 Y 通常是脯氨酸或羟脯氨酸残基。羟脯氨酸是由特定的酶催化，将脯氨酸残基做转录后修饰形成的。三股胶原蛋白链形成每三个残基一圈的螺旋，然后三股这种螺旋再形成超螺旋（见 2.1.4.6 节）。三螺旋中的链可以通过氢键相连从而形成上述的螺旋，这是因为胶原蛋白中无侧链的甘氨酸允许胶原蛋白链紧密排列（图 13.1）。

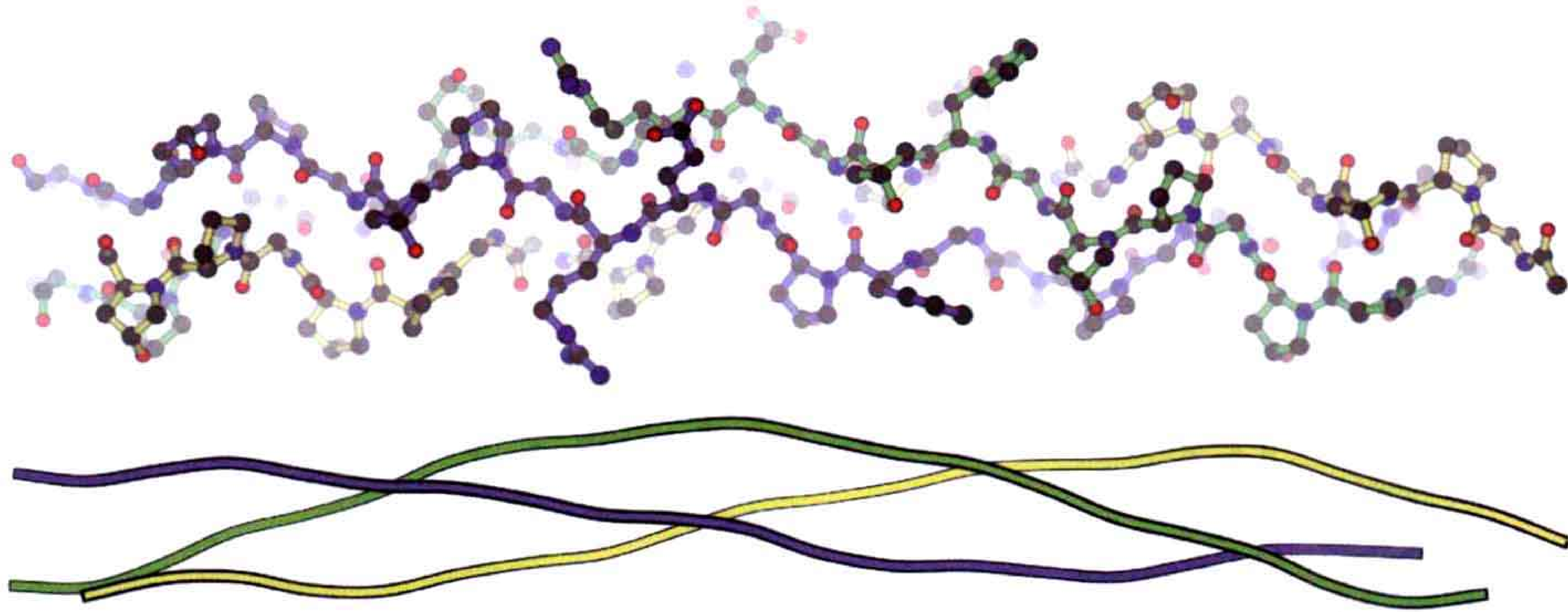


图 13.1　一段原胶原的三螺旋结构。上：球棍模型；下：绳子模型，其胶原蛋白链与上图使用了相同的颜色。三条链一级结构是脯氨酸 - 羟脯氨酸 - 甘氨酸的重复序列，其中甘氨酸残基允许链的紧密排列，形成链间氢键（PDB：1Q7D）。

13.1.2　纤连蛋白

细胞外基质含有纤连蛋白聚集体，细胞主要是通过与纤连蛋白的结合而贴附在细胞外基质上。纤连蛋白也存在于血浆中，以二硫键介导的二聚体形式存在。在细胞外基质中，每条纤连蛋白链结合几个其他分子，如胶原蛋白、硫酸乙酰肝素、血纤蛋白或整联蛋白。

纤连蛋白具有模块化的结构，每条链 220 kDa，含有 6 个结构区域，每个结构区域又是由很多模块组成的（图 13.2）。几种不同的纤连蛋白链都是由一个基因编码的，然后通过 mRNA 的选择性剪接形成了不同的纤连蛋白。组成纤连蛋白的模块分为 I 、II 和III型，不同的纤连蛋白分子中含有的模块种类和数量都不相同。根据序列的相似性，在其他的蛋白质中也发现了与纤连蛋白类似的模块。

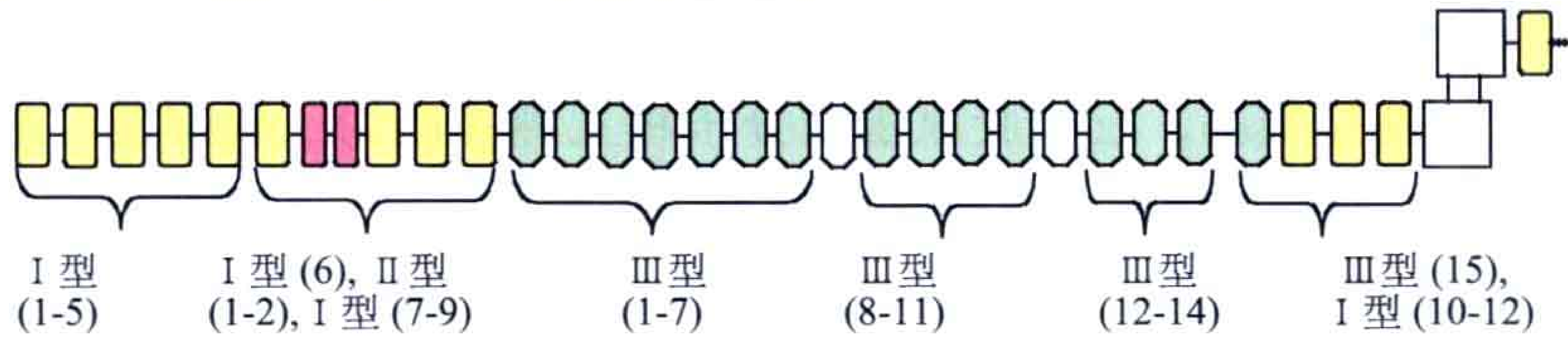


图 13.2　一个人类纤连蛋白单体的结构示意图。根据其结合蛋白的不同，纤连蛋白可以分成 6 个结构区域，结构区域 1 和结构区域 3 与其他纤连蛋白相互作用；结构区域 2 与胶原蛋白相互作用；结构区域 4 通过整联蛋白与细胞相互作用；结构区域 1 和结构区域 5 与肝素相互作用；结构区域 1 和结构区域 6 与血纤蛋白相互作用。每个结构域由不同数量的三种不同类型的模块组成，I 、II 、III分别用黄色、粉红色和绿色显示。这三种模块分别从 N 端编序号（见括弧中）。纤连蛋白链与链之间通过 C 端（白色方块显示）的二硫键相连。

I 型纤连蛋白模块含有 45 个氨基酸，构成了一种 β 发夹堆积在一个很小的 β 折叠片上的结构。由于这个结构很小，一般来说不太稳定，而该结构通过分子内二硫键连接这些二级结构，稳定了该结构模块。纤连蛋白的 N 端结构区域含有 5 个 I 型模块，对胞外基质的组装非常重要，但是目前还没有它与其他蛋白质相互作用的结构信息。

该结构区域中的第 4 和第 5 模块可以与血纤蛋白结合。第 4 和第 5 模块的相对位置固定，是因为第 4 模块的一个色氨酸结合在第 5 模块的口袋中（图 13.3）。

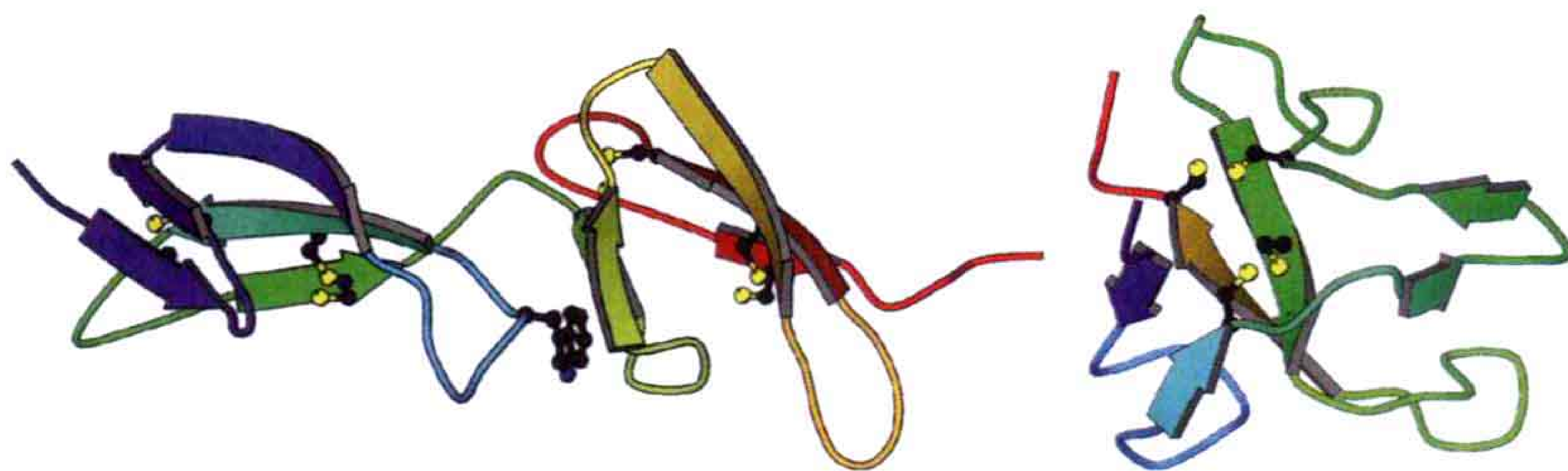


图 13.3 纤连蛋白 I 型和 II 型模块。左：人类纤连蛋白结构域中的第 4 和第 5 个 I 型模块。在 β1 与 β4 和 β4 与 β5 之间有二硫键连接，第 4 模块中的一个色氨酸结合在第 5 模块中的一个口袋中（PDB：1FBR）。右：II 型模块（PDB：2FN2）。

纤连蛋白中的模块和其他由模块组成的蛋白质都是小的但是独立存在的折叠单位。在大多数情况下，单个模块的亲和力较低，而高亲和力需要大面积的接触，这对于小结构单元是不可能的。因此，与其他分子的结合通常需要多个模块的共同参与，这意味着模块间的取向可能也需要被固定，如上述纤连蛋白 N 端结构域中的第 4 和第 5 模块的直接相互作用。

II 型模块也很小，为 45~60 个氨基酸残基，其结构与 I 型模块相似，同样是通过形成二硫键而被稳定，但其形成二硫键的位置与 I 型不同。

III 型模块是一种类似于免疫球蛋白的 β 三明治结构域，存在于 2% 的动物蛋白中。含有纤连蛋白 III 型模块的蛋白质还包括受体蛋白、胞外蛋白和胞内蛋白。这种模块参与了很多类型的相互作用。通常来说，连接 β 折叠链的环状区决定了相互作用的特异性。在纤连蛋白中，III 型模块参与其与整联蛋白和肝素的相互作用。

纤连蛋白与整联蛋白的结合主要通过第 10 个模块，该模块的一个环状区中含有一个 RGD 序列模体（图 13.4），该环状区比 III 型模块中的其他环状区都长。另外，与整联蛋白的结合也同时需要第 8 和第 9 模块及第 9 模块中的一个被称为“协同作用位点”（synergy）的参与。模块间的相对取向似乎对于结合整联蛋白非常重要。7~10 模块及其他一些包含 III 型模块片段的结构已经被解析，第 9 和第 10 模块相对取向大致一样，这是与众不同的。

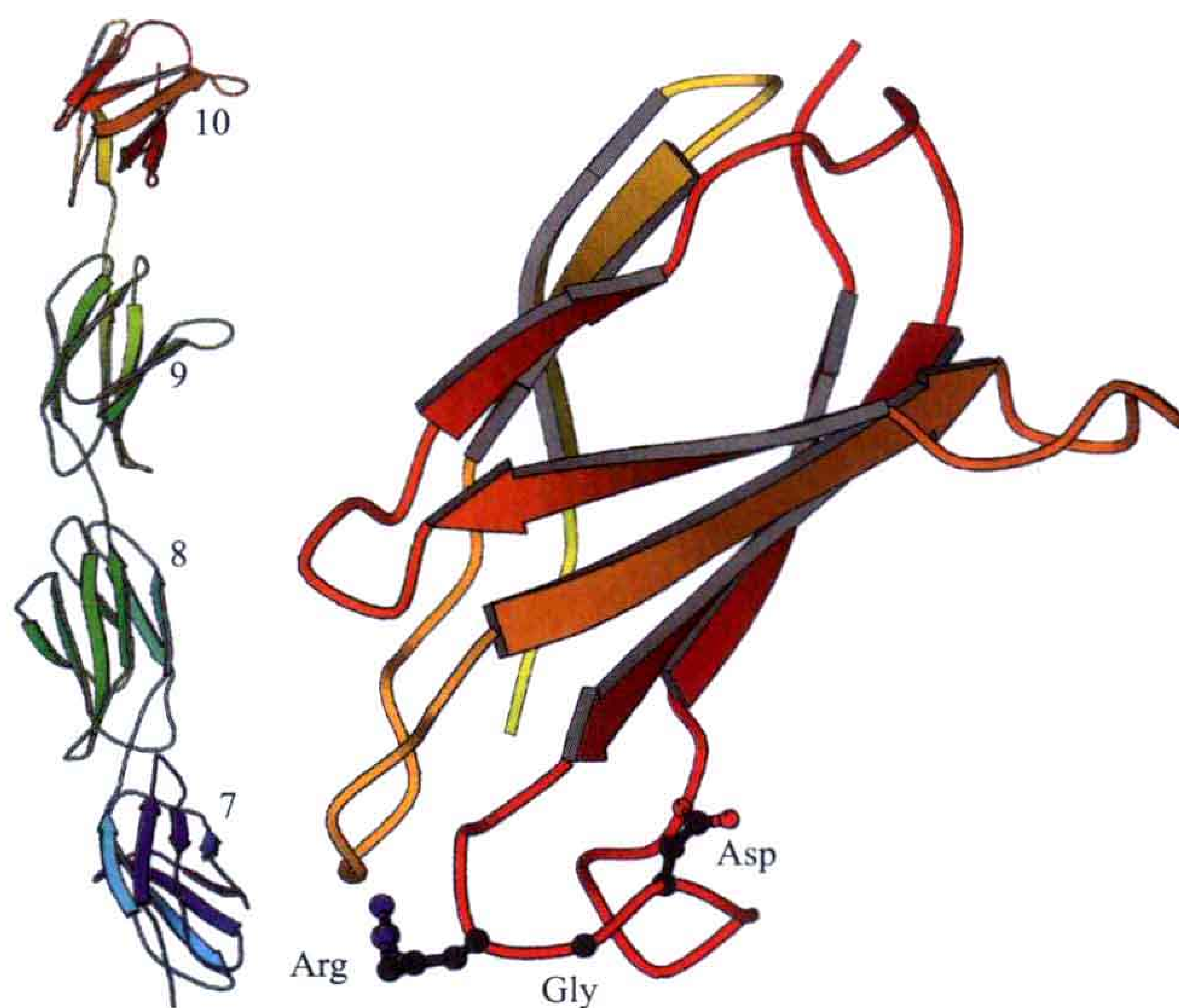


图 13.4 纤连蛋白 III 型模块。左：人类纤连蛋白 7~10 模块（PDB：1FNF）；右：第 10 模块。RGD 序列模体的残基的侧链以球棍模型显示（PDB: 1FNA）。

13.2 细胞与细胞外基质或其他细胞的连接

细胞含有大量的细胞外黏附分子（cellular adhesion molecule, CAM）参与细胞与细胞外基质或其他细胞的直接接触。主要的CAM家族见表13.1。本章中，只讨论整联蛋白家族。

表 13.1 细胞外黏附分子的不同家族

家族名称	接触类型	折叠结构
钙黏蛋白（cadherin）	同种细胞之间的接触	含有5个钙黏蛋白折叠的结构域，与其他细胞的钙黏蛋白结合
Ig类CAM	同种细胞之间的接触（原文如此，但不准确。有异种细胞之间Ig类CAM的接触，如CD2/CD58——译者）	几个Ig结构域，与其他细胞的CAM结合
整联蛋白	与细胞外基质的黏附	两个多结构域链
选择素	不同种细胞之间的接触	N端凝集素结构域（C型凝集素）。与其他细胞的碳水化合物结构结合

注：所有这些蛋白质都具有一个跨膜螺旋和一个小的胞内结构域。

13.2.1 整联蛋白的构成和结构

整联蛋白是一种连接在膜上的蛋白质，其伸在细胞外的部分与细胞外基质结合或参与细胞与细胞间的相互作用。整联蛋白可以介导细胞内向细胞外和细胞外向细胞内两个方向的信号传递。细胞内向细胞外的信号传递，活化信号来自细胞质，由其他受体分子激发，从而改变了整联蛋白的构象，使其可以高亲和地与细胞外配体结合。配体结合在整联蛋白的细胞外结构域，也会改变其构象，使胞内感受到来自细胞外的信号。

整联蛋白由α和β两条链组成，其中α链较长，为150~180 kDa；而β链较短，约90 kDa。它们是细胞表面受体，且α和β两条链都具有胞外部分、跨膜区和胞内大约50个残基的部分。在人类基因组中至少存在18种α链和8种β链，根据细胞类型的不同，它们可以形成多种组合（表13.2）。不同的整联蛋白结合不同的配体，如细胞外基质中的胶原蛋白和纤连蛋白，以及其他细胞的表面蛋白。整联蛋白通常识别的序列是精氨酸-甘氨酸-天冬氨酸（RGD）序列，这些序列存在于诸如纤连蛋白等的蛋白中。

表 13.2 一些整联蛋白及其配体

β链	α链	配体举例	识别序列
β1	α1、α2、α10、α11	胶原蛋白、层粘连蛋白	
	α4、α9	纤连蛋白、VCAM-1	LDV
	α5、α8、αV	纤连蛋白、波连蛋白（vitronectin）	RGD
	α3、α6、α7	层粘连蛋白	
β2	αL	ICAM、ICAM-2	
	αM、αX	补体C3b、纤连蛋白、ICAM	
β3	αV、αIIb	纤连蛋白和很多其他蛋白质	RGD
β4	α6	层粘连蛋白	

续表

β 链	α 链	配体举例	识别序列
β5	αV	波连蛋白	RGD
β6	αV	纤连蛋白	RGD
β7	α4	纤连蛋白、VCAM-1、MAdCAM-1	LDV

注：β2 整联蛋白的 α 链和 α1、α2、α10 和 α11 的头部区域含有一个插入结构域。

整联蛋白的胞外部分具有一个由 α 和 β 亚基的 N 端结构域构成的头部。头部与跨膜区通过一个杆部相连（图 13.5）。整联蛋白 αVβ3 的大部分胞外结构已被解析。β 链的头部含有三个结构域：I 结构域或 A 结构域、hybrid 结构域和 PSI（plexin-semaphorin-integrin）结构域（图 13.6）。hybrid 结构域插入 PSI 结构域中；而 I 结构域则插入 hybrid 结构域中。hybrid 结构域属于一种特别的 Ig 折叠。整联蛋白 αVβ3 中的 PSI 结构域是无序的，但可以从 β 链片段的结构中获知。I 结构域负责与配体结合，它具有一个类似于很多核酸结合蛋白中存在的 Rossmann 折叠的结构。

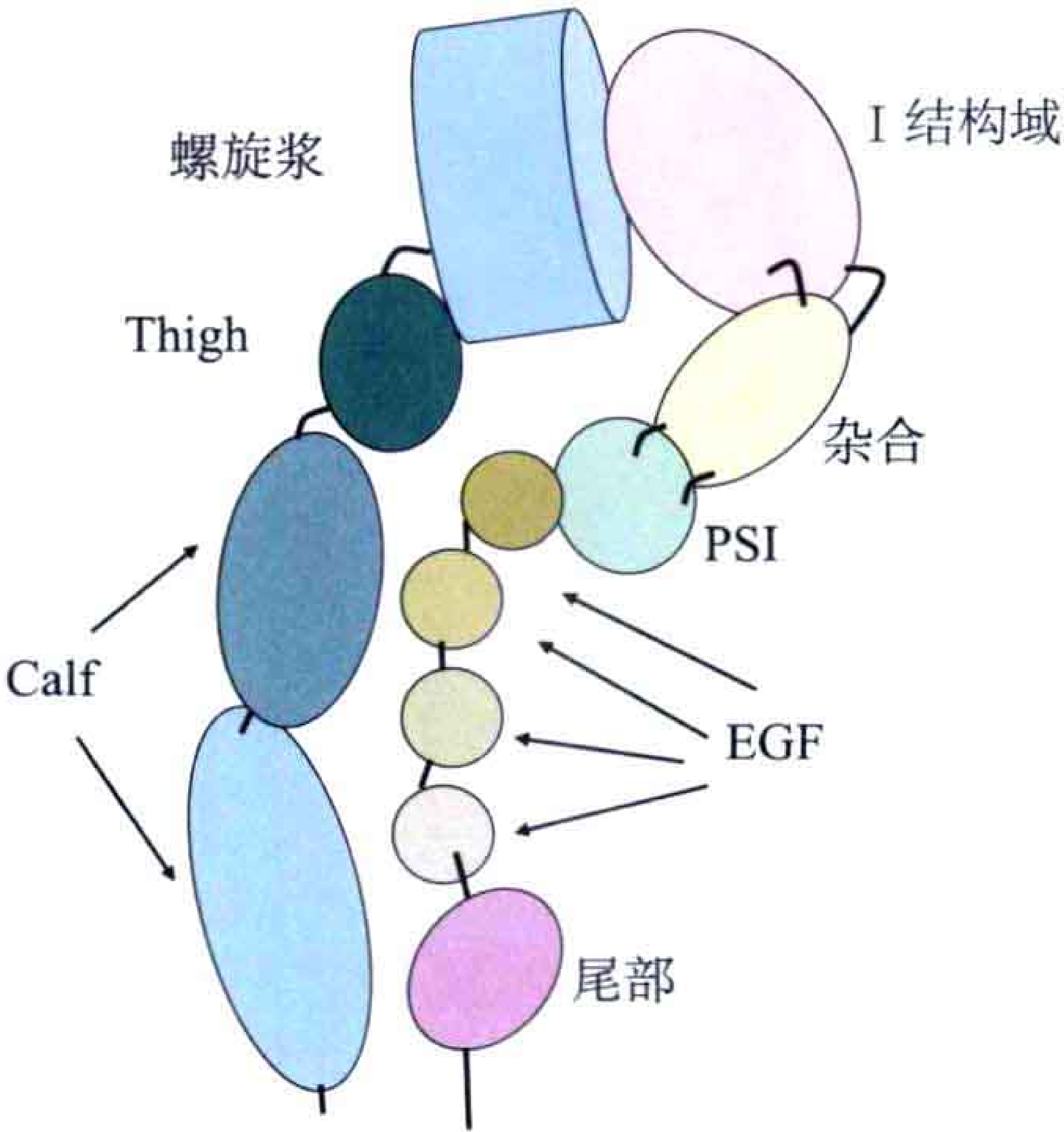


图 13.5　整联蛋白的胞外部分的示意图。

α 链的头部含有一个与三聚 G 蛋白的 β 链相同的螺旋桨结构域（图 13.6）。螺旋桨结构域与 I 结构域之间的相互作用也与三聚 G 蛋白 α 和 β 亚基的相互作用类似。I 结构域中的 Rossmann 折叠也与 G 蛋白的结构类似，但是中心 β 片层 N 端部分的 β 折叠链之间的连接方式不同。这种相似性表明螺旋桨结构域与 I 结构域之间通过类似于三聚 G 蛋白的螺旋桨结构域阻断 α 亚基的方式，来调节整联蛋白的活性。

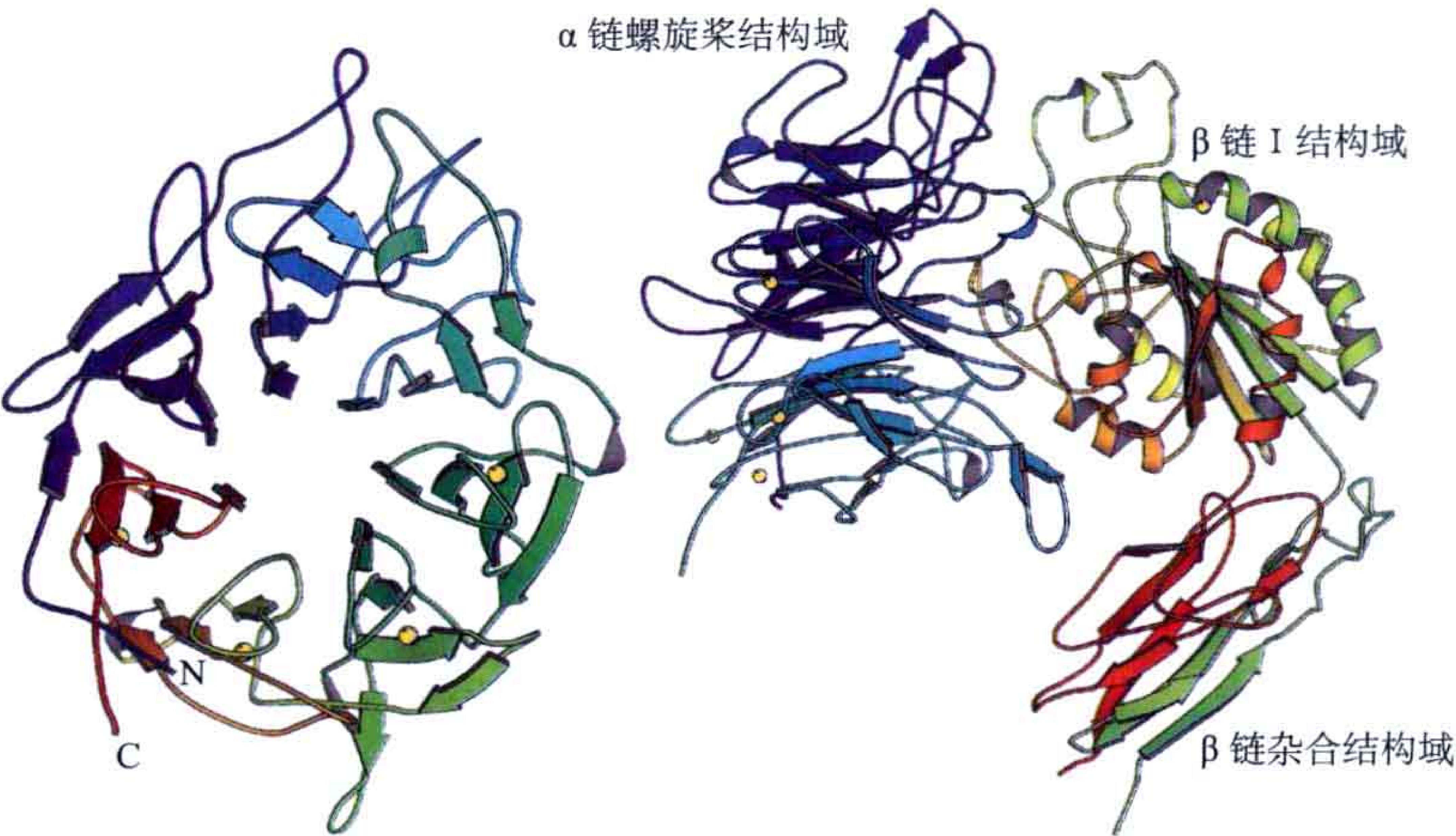


图 13.6　左：αVβ3 整联蛋白 α 链中的 β 螺旋桨结构域。黄色球显示的是 4 个钙离子结合位点。它们结合在 7 个螺旋片中的 4 个相应位置。右：胞外部分的头部。β 链的 I 结构域与 α 链的 β 螺旋桨结构域（α 螺旋桨）相互作用。

整联蛋白杆部包括 α 链的多个 Ig 类型结构域和 β 链的多个 EGF 结构域（图 13.7）。β 链的 C 端结构域具有一个独特的折叠，通常被称为 β- 尾结构域。

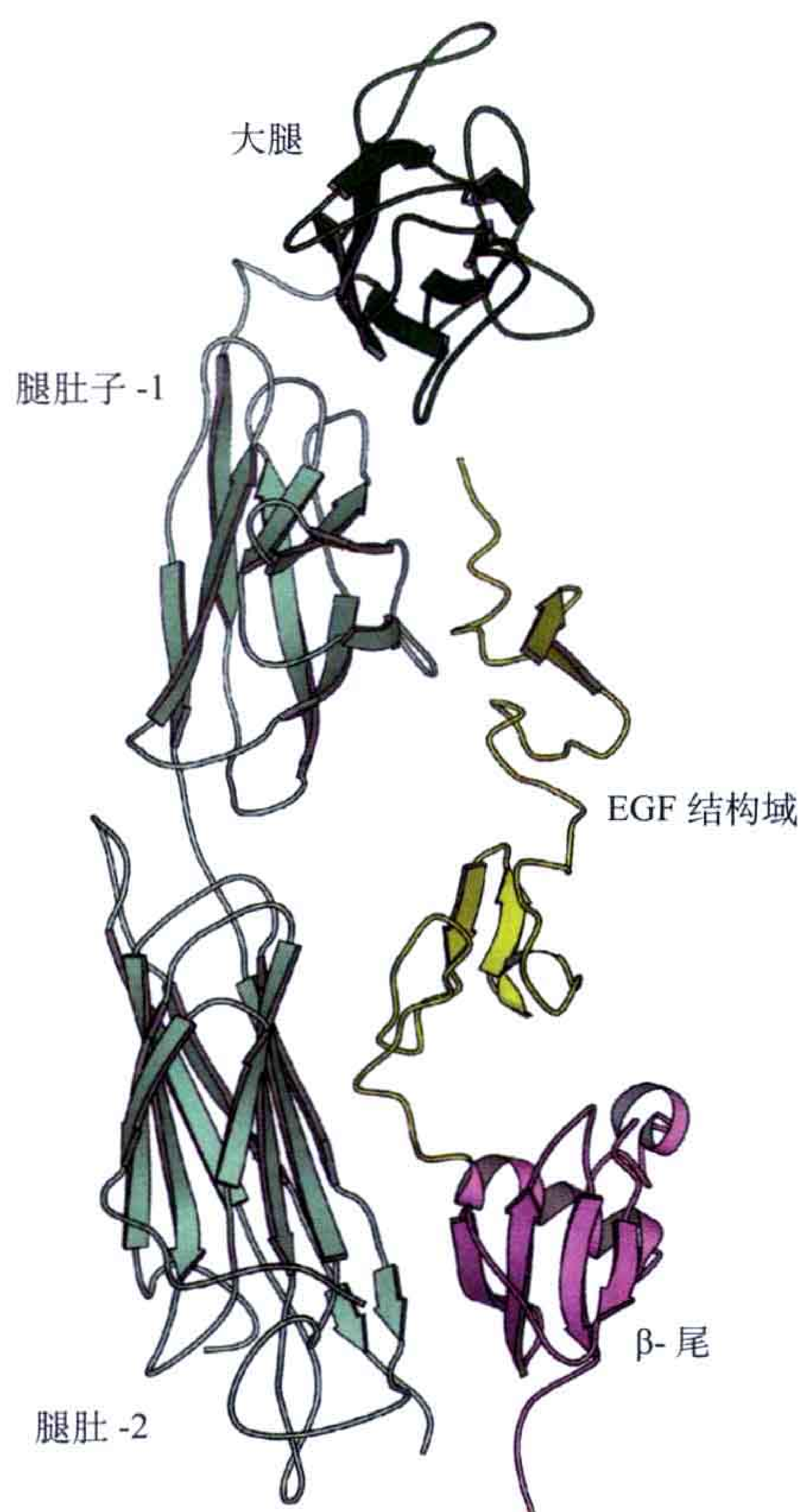


图 13.7　整联蛋白 αVβ3 的杆部。β 链中只有两个 EGF 结构域在结构中可见。

很多整联蛋白在其 α 链的螺旋桨结构域中含有插入结构域，形成与 β 链的 I 结构域相同的另一个 I 结构域。β2 整联蛋白和一些其他整联蛋白中的 α 链就含有插入结构域（表 13.2）。整联蛋白若含有 α I 结构域，则该蛋白会以这个结构域来与配体结合。

13.2.1.1　MIDAS 模体

整联蛋白的 I 结构域以一种由金属离子介导的独特方式与其配体结合。两个丝氨酸和一个天冬氨酸的侧链参与和金属离子形成配位键（图 13.8）。金属离子诸如 Mg^{2+} 可以与 6 个氧原子配位，其中包括两个带负电荷的氧原子。I 结构域中的金属离子与该结构域内氧原子的配位并不完全，所以其他蛋白质可以通过带负电荷的氨基酸残基（天冬氨酸或谷氨酸）与金属离子结合。水分子与金属离子的配位结合进一步使其形成完全配位的状态。整联蛋白中这种结构被称为 MIDAS（metal-ion dependent adhesion site）模体。

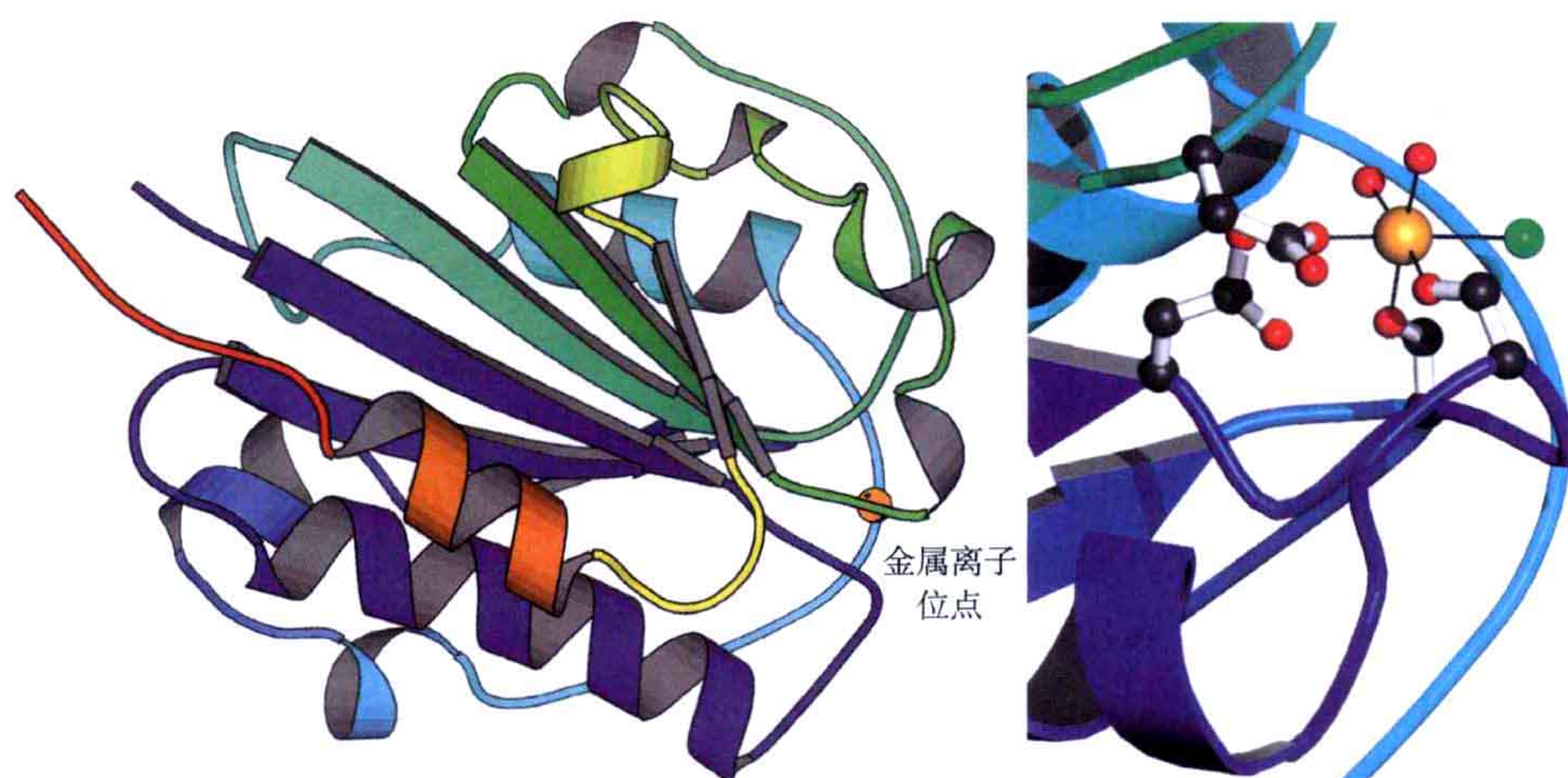


图 13.8　左：αL（CD11a）中的 I 结构域示意图；右：CD11a 中的金属结合位点的放大图。金属离子（橙色，在该晶体结构中是锰离子，虽然体内该酶结合的是镁离子）配位结合两个丝氨酸（下方）和一个天冬氨酸（左）。在该晶体结构中，两个水分子（红色）和一个氯离子（绿色）也参与了与金属离子的配位结合，形成 6 配位。这是 MIDAS 关闭状态的构象（PDB：1LFA）。

I 结构域与一段类似胶原蛋白结构的肽段的复合体结构证实了金属离子与酸性氨基酸残基的相互作用对结合是非常重要的。这个含有 21 个氨基酸残基的肽段具有类似胶原蛋白的序列，其形成了三螺旋结构，并通过一个谷氨酸残基的侧链与整联蛋白的 MIDAS 结合，结合表面的面积较大，而且同时具有极性和非极性的相互作用（图 13.9）。

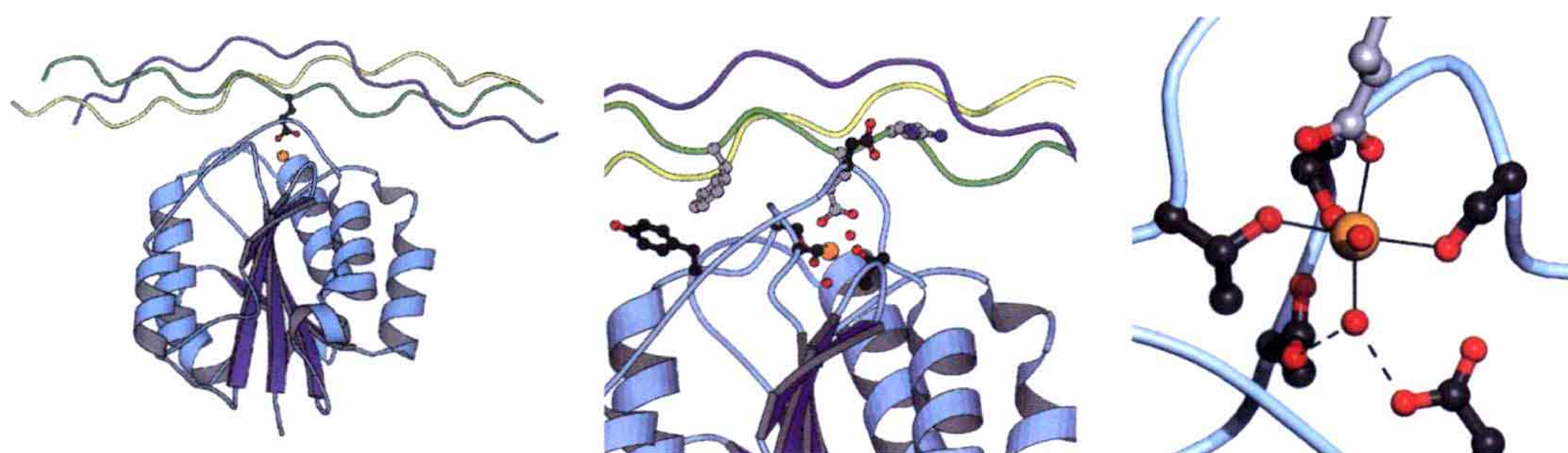


图 13.9　胶原蛋白与整联蛋白 α2β1 的 I 结构域的结合（PDB：1DZI）。上：胶原蛋白三螺旋与 I 结构域的结合。左下：金属（橙色，在该晶体结构中是金属钴离子）与氧原子配体结合位点的放大图。一些参与分子间直接相互作用的氨基酸用球棍模型显示，胶原蛋白中的侧链碳原子显示为灰色。右下：金属配位的细节图。整联蛋白中有分别来自两个丝氨酸和一个苏氨酸（黑色）的三个氧原子与金属的配位，两个水分子和胶原蛋白中的一个谷氨酸（灰色）侧链氧原子完成了金属原子的 6 配位（配位键用实线表示）。整联蛋白中的两个天冬氨酸与和金属配位的水分子之一形成了氢键（虚线表示）。这是 MIDAS 模体的开放、高亲和力的构象。

13.2.2 调控整联蛋白结合的机制

整联蛋白以不同的亲和力与配体结合。这是一个调节信号传递的机制，胞外信号向胞内传递时，配体的结合影响胞内的细胞骨架；胞内信号向胞外传递时，细胞的状态影响整联蛋白的亲和力。目前对整联蛋白的亲和力调控机制已有了部分了解。

αⅤβ3 胞外部分的晶体结构呈现一种弯曲的构象（图 13.10）。电镜结构显示，整联蛋白的胞外部分存在三种不同的构象。αⅤβ3 显示的弯曲构象是一种非活化的状态。另外两种较为伸展的形式，被称为关闭和开放的状态。这些状态的区别在于其跨膜区的相对位置，开放的构象对应活化的状态。我们已经知道了一个处于活性状态的头部的晶体结构（整联蛋白 αⅡbβ3 的头部结构），其与非活性状态的最大区别是 βⅠ结构域和 hybrid 结构域之间的角度（图 13.11）。在非活化的 αⅤβ3 结构中，α 和 β 两条链是紧靠在一起的，上述两个结构域之间角度的改变将会使两条链分离而形成活性状态的 αⅡbβ3 中的构象。

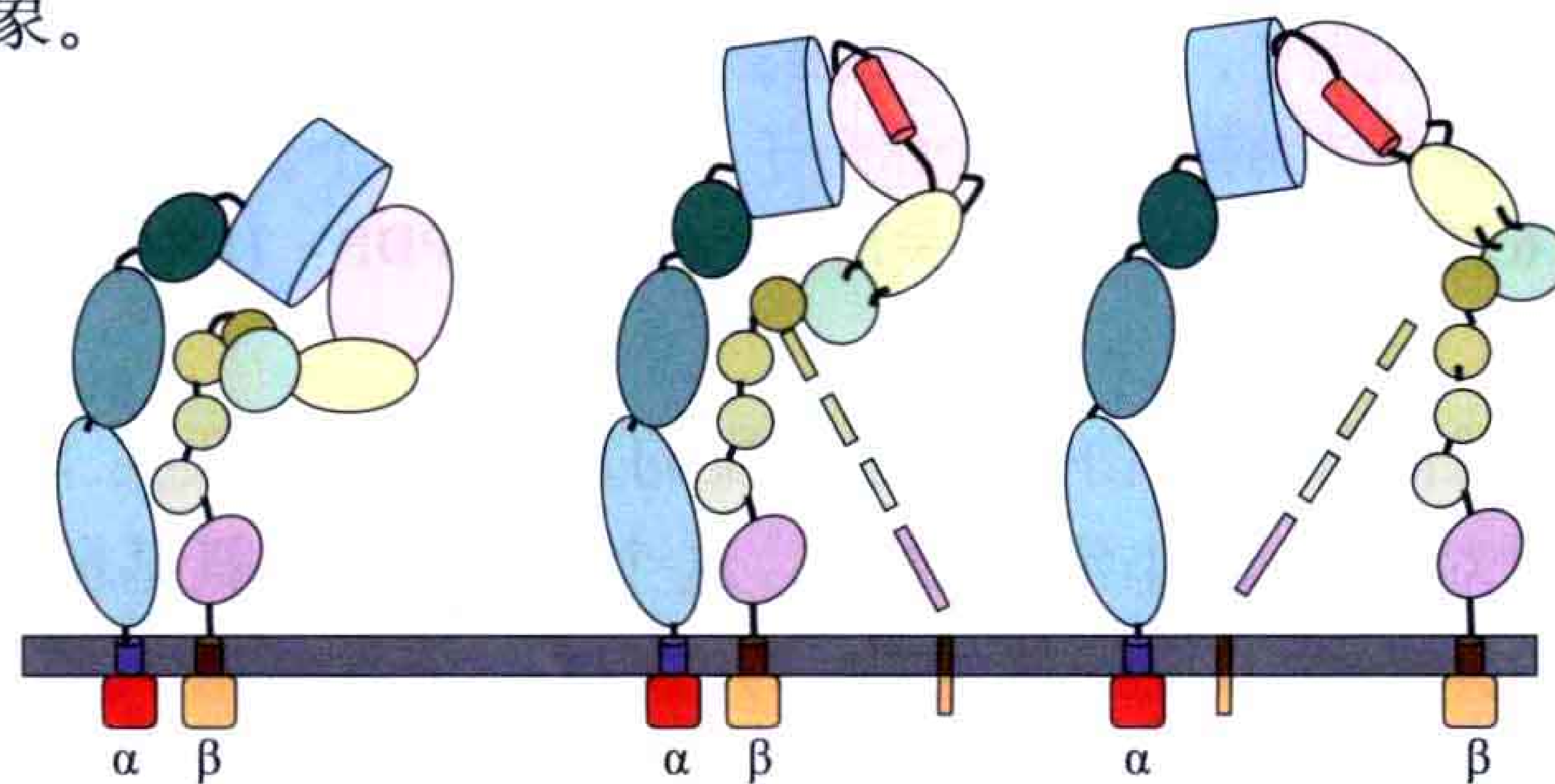


图 13.10　整联蛋白头部可能构象的示意图。弯曲（左）和关闭（中）状态的亲和力低，而开放状态（右）的亲和力高 [本图基于 Luo BH, Carman CV, Springer TA. (2007) Structural basis of integrin regulation and signaling. *Annu Rev Immunol* **25:** 619-647 修改而成]。

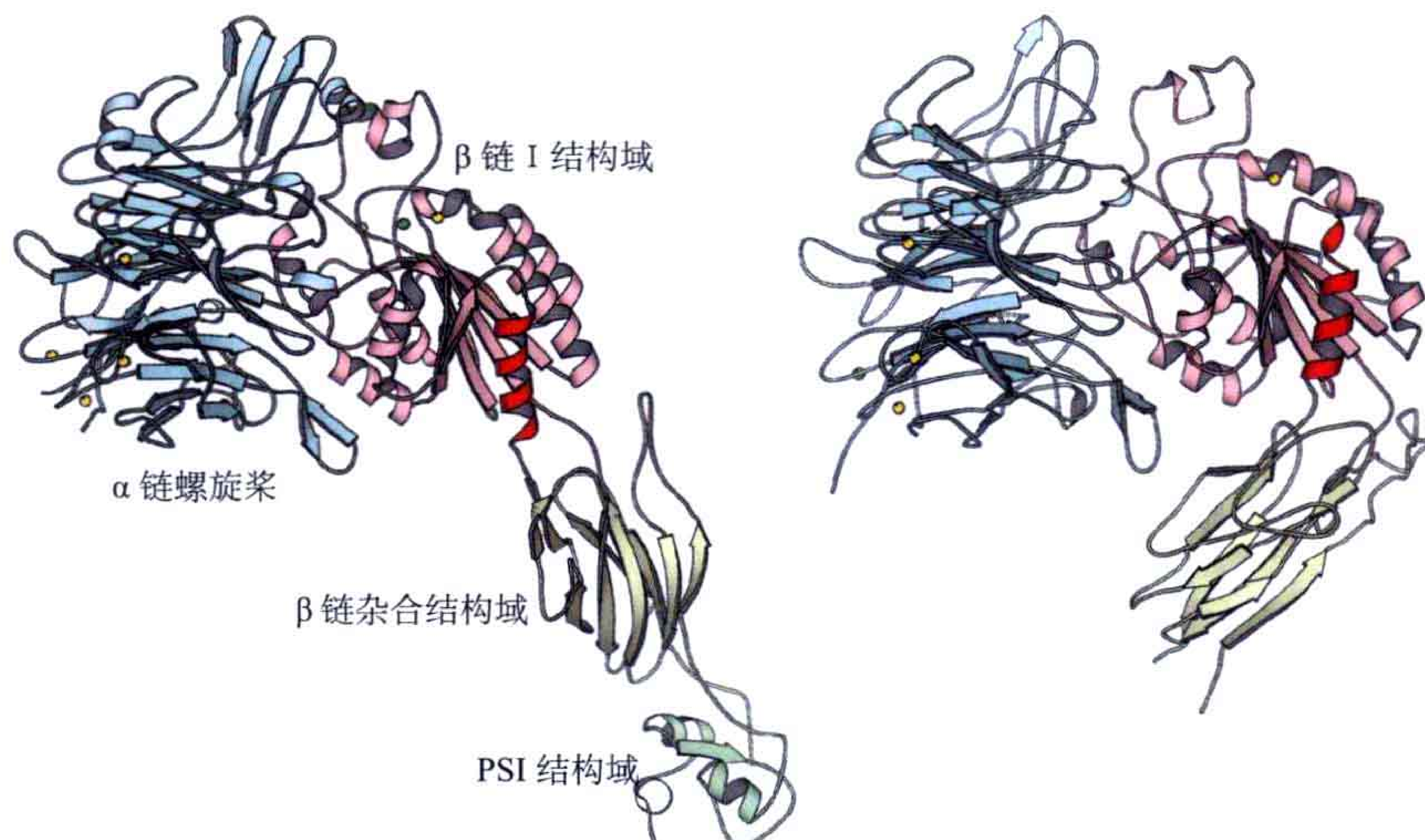


图 13.11　整联蛋白 αⅡbβ3（左，PDB: 1TY3）和 αⅤβ3（右，PDB: 1JV2）的头部以近似的取向显示。需要注意的是，Ⅰ结构域和 hybrid 结构域间的角度在这两个蛋白质中是不同的。该角度受到Ⅰ结构域（红色）C 端螺旋的调控。PSI 结构域只在 αⅡbβ3 结构中是有序的，因此在 αⅤβ3 结构中没有画出。

I 结构域低亲和力的关闭状态和高亲和力的开放状态（结合 MIDAS 模体的状态）的结构都已被解析，其主要的区别是 C 端螺旋（α7）发生了大约 9 Å 的位移（图 13.12）。这个位移与 MIDAS 中金属离子的配位偶联，而这种偶联使得整联蛋白可以将底物结合转化成为自身构象的变化。C 端螺旋的移动改变了 I 结构域与 hybrid 结构域之间的取向，这与 PSI 结构域末端大约 70Å 的位移相应（图 13.11）。

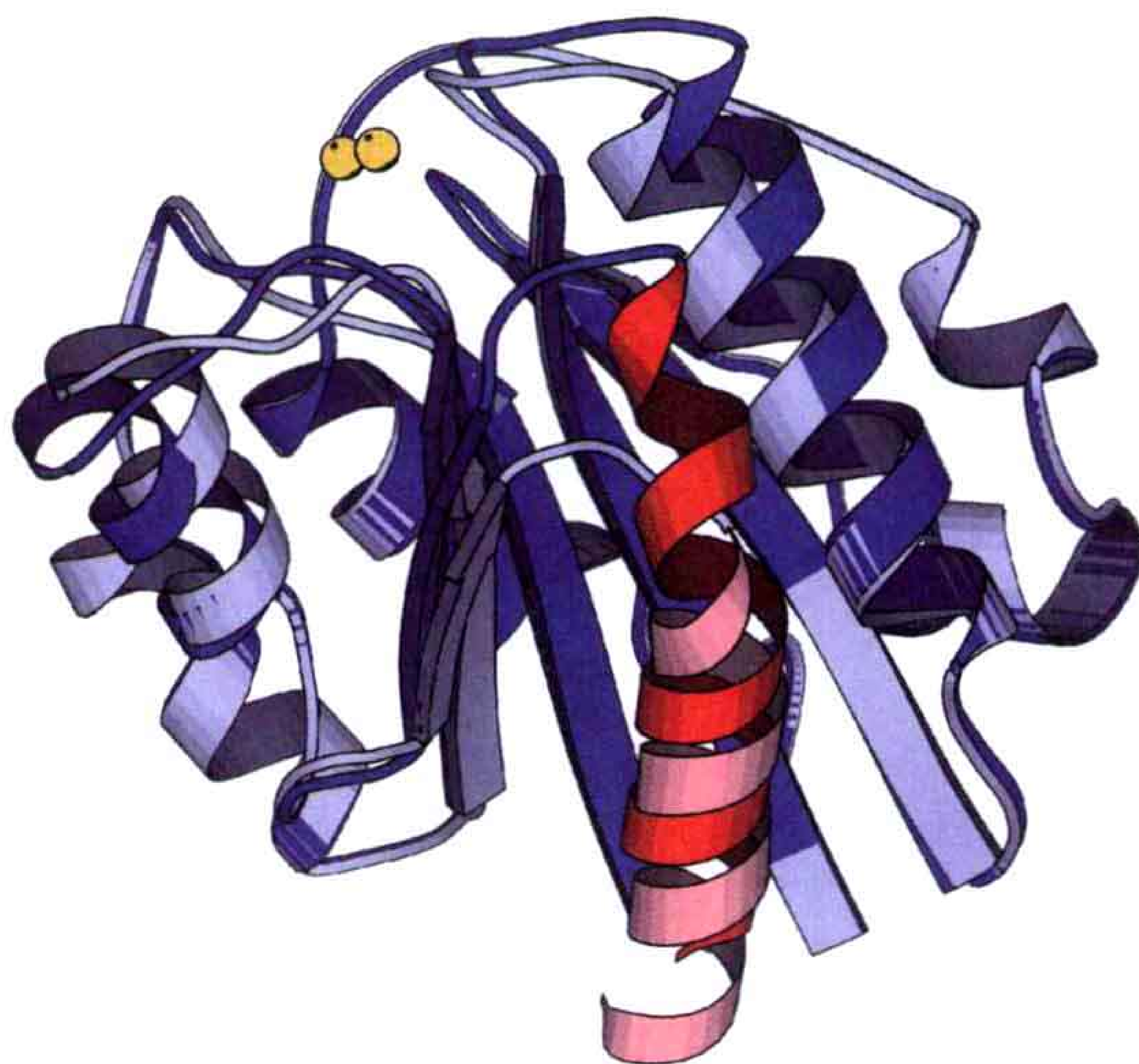


图 13.12　I 结构域开放（浅蓝色）和关闭（深蓝色）的状态。MIDAS 模体中的金属离子用黄色显示。这两种状态的主要区别是在蛋白质活化时其 C 端螺旋的移动。

具有两个 I 结构域的整联蛋白，其 α 链的 I 结构域（α I 结构域）负责结合底物。在这些整联蛋白中，α I 结构域和配体的亲和力与 β 链构象变化的偶联一定是非直接的，但是 α I 结构域中的一个带负电荷的氨基酸与 β 链 MIDAS 模体结合可能参与这个构象变化。如果这种结合调控 α I 结构域的亲和力，细胞内的状况和整联蛋白分子的整体构象则也可以调控这些整联蛋白的活性。

延伸阅读

原始文献

Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. (2000) Structural basis of collagen recognition by integrin α2β1. *Cell* **101**: 47-56.

Lee JO, Rieu P, Arnaout MA, Liddington R. (1995) Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* **80**: 631-638.

Xiao T, Takagi J, Coller BS, Wang JH, Springer TA. (2004) Structural basis for allostery

in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* **432**: 59-67.

Xiong JP, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL, Joachimiak A, Goodman SL, Arnaout MA. (2001) Crystal structure of the extracellular segment of integrin αVβ 3. *Science* **294**: 339-345.

综述文章

Luo BH, Carman CV, Springer TA. (2007) Structural basis of integrin regulation and signaling. Annu Rev Immunol 25: 619-647.

（王晓君 译　王家槐 校）

第十四章
免疫系统

所有的生物体必须保护自己免受病毒和其他生物的入侵，细胞也需要把异己与自身相区别。这个复杂的任务需要高度精密的系统来完成。所有的真核生物都拥有一个非获得性免疫系统 (non-adaptive system)，即先天性免疫 (innate immunity)。同时，脊椎动物还存在着一个获得性免疫系统。获得性免疫通过以下两条途径来抵御外来生物的入侵：抗体介导的免疫系统（antibody-mediated system），又称为体液免疫（humoral immunity）；T 细胞介导的免疫系统（T-cell mediated system），即细胞免疫（cellular immunity）。获得性免疫有两个关键特点：一是特异性强，这样就可以正确地识别外来分子以将其消灭；另外，它拥有一个记忆系统，当有"入侵史"的分子再度入侵时，组织器官可以迅速对其做出免疫应答。抗体 (antibody) 由 B 细胞产生，可以识别抗原 (antigen) 的表面决定簇（epitope）。体液免疫处理细胞外的病毒和细菌的感染。而细胞免疫对被病毒、细菌，以及其他任何类型的外源物质和组织感染的细胞做出应答。细胞表面的主要组织相容性复合体（MHC）分子将外源蛋白的肽段呈递给 T 细胞受体，那些含有外源物质的细胞就会随之被消灭。MHC, 又被称为移植抗原（transplantation antigen），因为这些分子非常多样化，任何两个个体不大可能有相同的一组 MHC 分子。移植细胞也正是由于其 MHC 分子与移植受体细胞的 MHC 分子间的不同而被辨别区分开的。如今由于大量免疫系统蛋白分子结构信息的获得，使得我们可以从结构细节上来理解免疫系统中那些发挥中心作用的重要分子间的相互作用。

14.1 体液免疫——抗体介导的免疫系统

在体液免疫中，抗体对于某个抗原的特异性识别是通过细胞的吞噬作用（phagocytosis）或激活补体系统（complement system）来实现的，并导致外源物质被中和或破坏。吞噬作用是指白细胞对外源物质进行消化。补体系统包括一系列可发生级联激活反应的丝氨酸蛋白酶，这种级联反应最终导致了补体系统通过裂解细胞膜而消灭外源细胞。

由于外源分子的海量数目，B 细胞产生的抗体或者免疫球蛋白需要高度的多样性才能达到特异识别特定抗原这一要求。而这又是如何实现的呢？为了回答这个问题，我们首先需要描述抗体分子构成。抗体是由相同类型的结构域重复构成的，这个结构域是一个 β 三明治结构域，含有两层反向平行的 β 折叠链（图 14.1）。

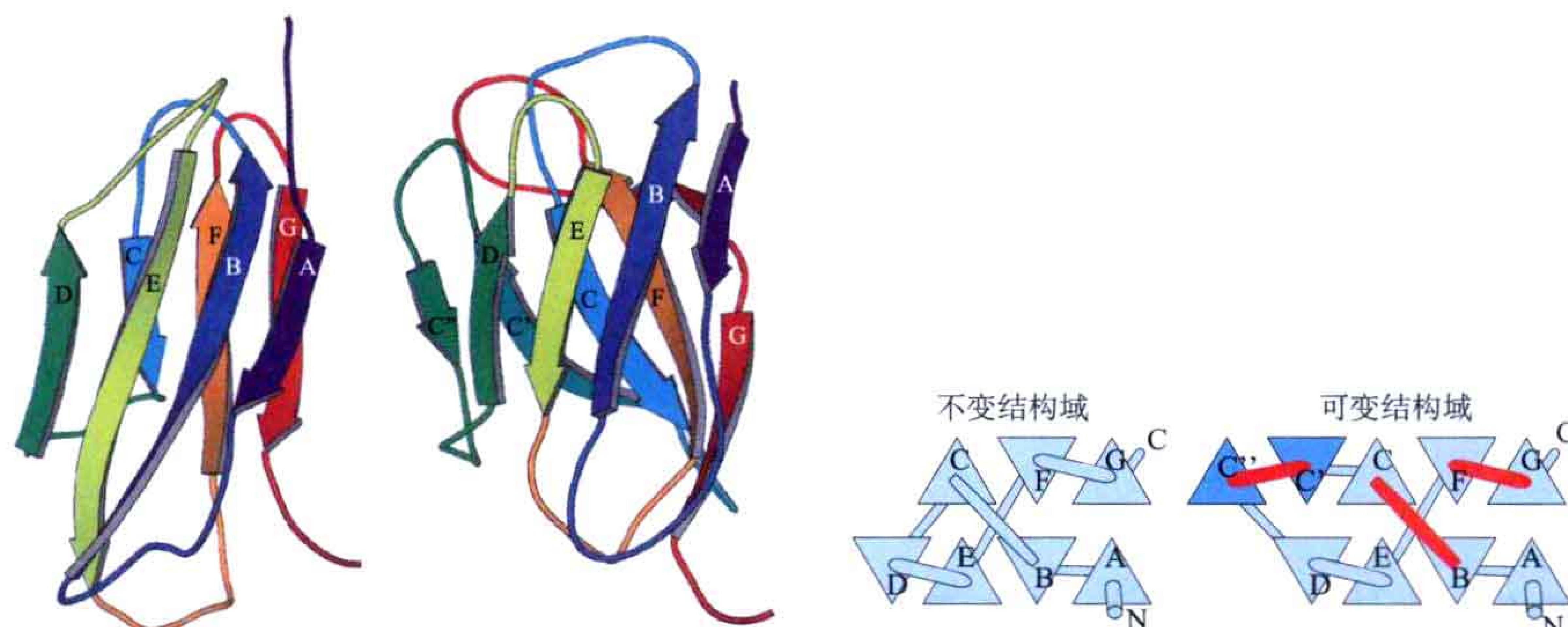


图 14.1 免疫球蛋白（immunoglobulin, Ig）折叠。免疫系统中的大量蛋白质都含有该折叠模式，细胞黏附和神经系统中的一些蛋白质也含有该折叠。左：一个不变结构域（constant domain）和可变结构域 (variable domain) 折叠方式的带状卡通示意图（PDB：1AQK，重链）；右：一个含有免疫球蛋白折叠的 β 三明治结构域的示意图。不变结构域折叠含有一个由 4 条反平行链及一个 3 条反平行链组成的两个 β 折叠层。可变结构域折叠比不变结构域多两个 β 折叠链，C′ 与 C″ （深蓝色）。其中红色的连线部分代表连接可变结构域内沿多肽链的一些 β 链的互补决定区（complementary-determining region），即 CDR1、CDR2、CDR3。这些区域形成抗原结合表面。

14.1.1 免疫球蛋白 G 分子

抗体由两种肽链构成，即轻链和重链。轻链通常包含两个结构域，每个大概含有 110 个残基。轻链主要包括两类：κ 和 λ。重链至少含有 4 个结构域。在哺乳动物中，有 5 种不同的抗体类型，即 IgA、IgD、IgE、IgG 和 IgM，它们的功能及器官分布都不同。与哺乳动物相比，其他脊椎动物在抗体的构成方式上比较有限。所有抗体都含有不同类型的重链，可以形成不同的多聚体。在血浆中，IgG 分子是最常见的免疫球蛋白。它含有两条重链和两条轻链，共 12 个免疫球蛋白结构域（Ig 结构域）（图 14.2）。重

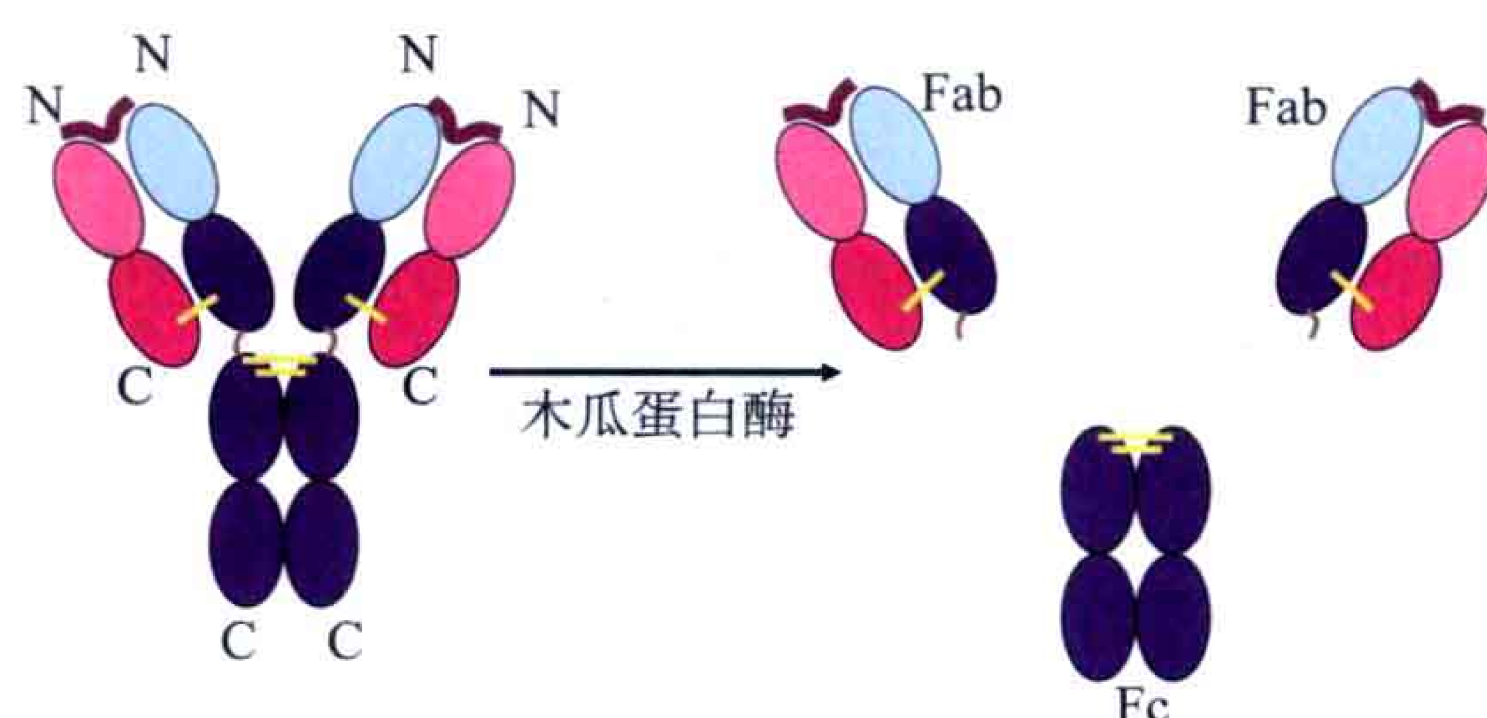


图 14.2 IgG 分子由 4 条多结构域肽链构成，两个重链（蓝色）和两个轻链（红色）。重链由 4 个结构域构成，而轻链由 2 个结构域构成。浅蓝色和浅红色所示的结构域是可变结构域，该处含有抗原结合位点。深色所示的结构域为不变区。重链之间由二硫键连接，每条轻链与一条重链由二硫键链接，黄色连线表示二硫键。如果用蛋白酶如木瓜蛋白酶处理 IgG 分子，重链在第一和第二不变结构域间被酶切断开，这样整个分子就分成了三个片段。只含有重链不变结构域的片段成为 Fc 片段，Fc 代表可结晶片段（fragment crystalline，早年刚获得这个片段时，它在透析袋内自动长出晶体来）。另外两个一样的片段被称为 Fab 片段（fragment antigen-binding，抗原结合片段）。抗原（紫色）结合在 Fab 片段上的两个可变结构域之间。

链和轻链氨基端的结构域配对形成了抗原结合结构域。一个 IgG 分子含有两个完全相同的位点用于结合抗原。

Ig 结构域通常以成对的方式起作用。在 Fab 片段上的重链和轻链形成的结构域中，它们以异质对（hetero-pair）的方式结合；而在 Fc 片段中，以同质对 (homo-pair) 方式结合（图 14.2 和图 14.3）。此外抗体也存在由轻链形成的同源二聚体，如由特定的癌细胞大量产生的被称为 Bence-Jones 的蛋白。

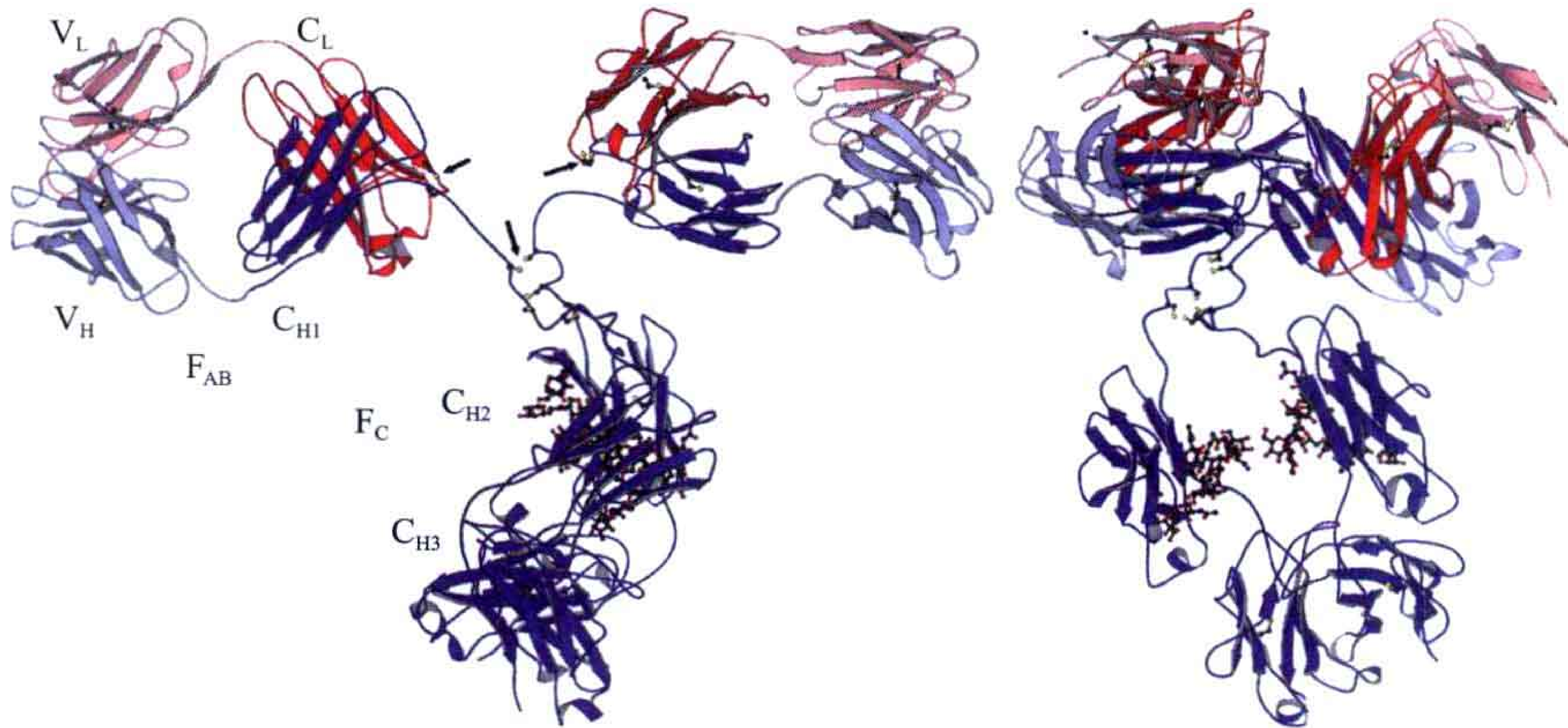


图 14.3　两个方向的 IgG 分子结构细节图。重链包含 V_H、C_{H1}、C_{H2} 和 C_{H3} 结构域，轻链包括 V_L 和 C_L 结构域。图中，处于上方的 Fab 片段与下方的 Fc 片段通过柔性的部分连接。重链间及重链与轻链间由二硫键连接。所有参与二硫键形成的半胱氨酸全部显示了出来，参与链与链间连接的二硫键用箭头加以标识。Fc 片段在 C_{H2} 结构域上的糖基化修饰以球棍模型的形式显示（PDB：1IGT）。

与抗原结合的 4 条链的氨基端部分也称为可变结构域（V_H 或 V_L）。羧基端的结构域被称为不变结构域（C_H 或 C_L）。不变结构域通常由含 7 个 β 折叠链的两层三明治结构构成，而可变结构域含有两个额外的 β 折叠链（图 14.1）。

14.1.2　抗原的识别

抗体可以与数以百万计的不同抗原进行特异结合。在几天或几周的时间内，一个脊椎动物可以针对一个新的外源分子产生大量的抗体。重链和轻链上的互补决定区（CDR1、CDR2 和 CDR3，图 14.1）拥有巨大的识别能力，其负责与不同的抗原识别和结合的区域。每个抗原结合表面中 6 个高度可变的环状互补决定区形成一个用于结合抗原的、连续的、大约 1000Å^2 的表面。该表面由较平坦的部分以及沟槽和缝隙构成用于与抗原的特异性结合。

抗体可以和被称为半抗原（hapten）的小分子结合。抗体与半抗原、肽段和蛋白质的结合已经成功从结构上得以描述。抗体与抗原蛋白的相互作用面首先是在溶菌酶及流感病毒的神经氨酸苷酶与它们各自的单克隆抗体的晶体学复合物结构中得以分析的（图 14.4）。抗体的结合表面包含大部分或全部的互补决定区，这些互补决定区形成较为平坦的区域，在形状和表面电势上和与之结合的抗原分子正好匹配。复合物，尤其是溶菌酶与抗体的复合物中所包含的单个蛋白的结构已经得以解析，通过单个蛋白结构与复合

物中蛋白质的结构的比对发现，蛋白质在形成复合物时发生的构象变化很小。

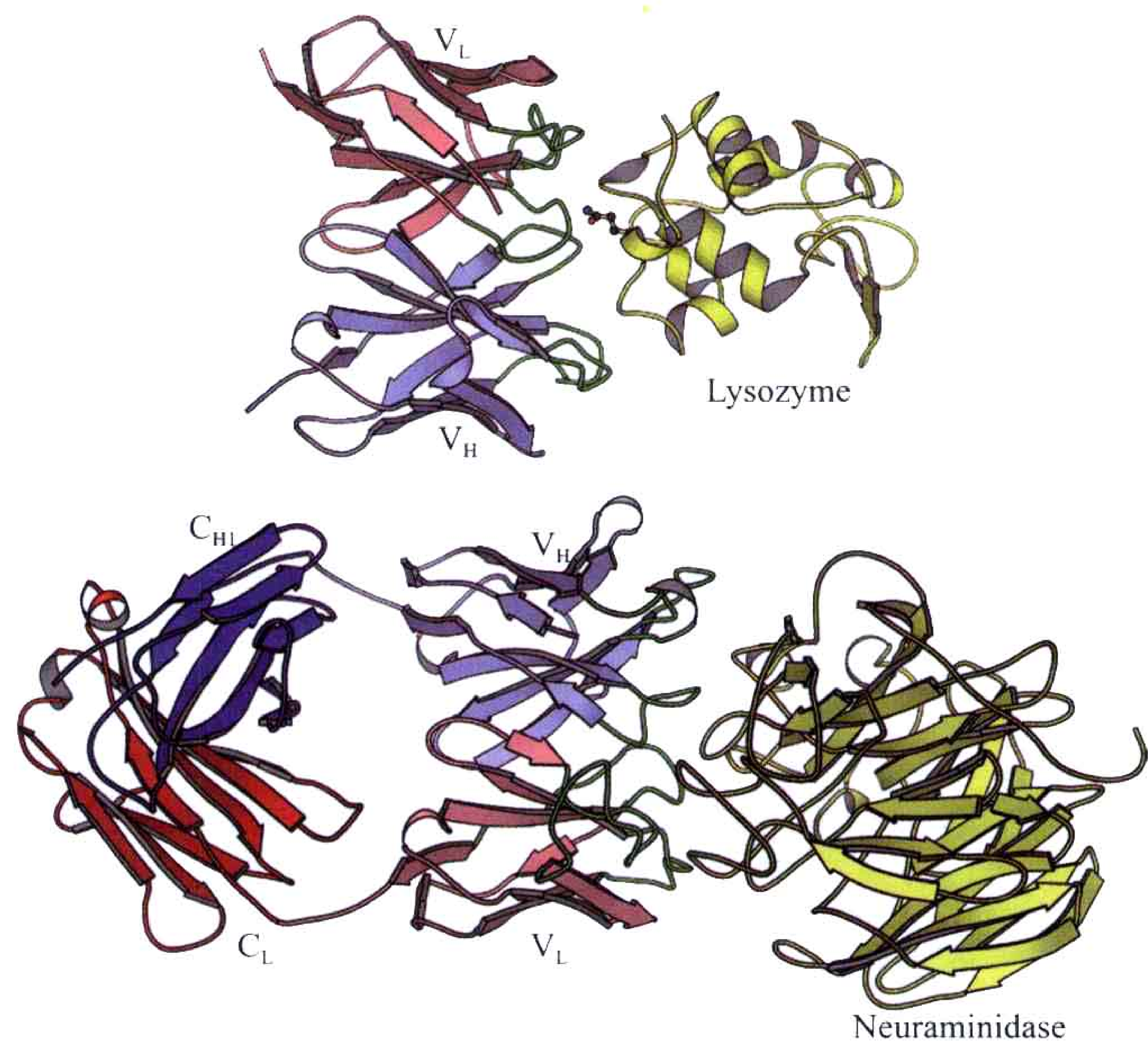


图 14.4　抗原与其单克隆抗体间的结合。上：Fab 片段可变结构域结合到一个溶菌酶分子上（PDB: 1VFB）；下：Fab 片段结合到来自流感病毒的神经氨酸苷酶上（PDB: 1NCA）。高变环区（即 CDR 区域）用绿色显示。

由有限的基因构成的基因组是如何产生成千上万种抗体的呢？例如，人类基因组大概一共有 30 000 个基因。基因重组是造成抗体多样性的原因。人类基因组中含有很多基因片段，它们以不同方式进行重组由此产生抗体。轻链（κ 或 λ）可以由三种不同的基因片段以任意结合方式重组后表达产生。这三种基因片段包括：V（variable, 可变）片段（κ 和 λ 的拷贝数都大于 70）、J（joining, 连接）片段（κ 含有 5 个拷贝，而 λ 至少含有 7 个拷贝）、C（constant, 不变）片段（κ 只含有 1 个拷贝，而 λ 至少含有 7 个拷贝）（图 14.5）。重链含有多于 100 个拷贝的 V 片段、9 个 J 片段，11 个 C 片段，以及在轻链中所没有的 D（diversity, 多样化的）片段。然而这其中的一些片段是很少或根本不会被用到的。轻链中的 J 片段，重链中的 J 和 D 片段编码 CDR3 的一部分。这就提高了 CDR3 的高变性特点，尤其是重链的高变性。基因的连接也不总是正好在同一个位置，这也进一步增加了高变环区的可变性。除此之外，当 B 细胞成熟时，CDR 区域的高频率突变也随之发生，这也使得重链和轻链的可变性概率大为增加。而重链和轻链间的随机配对将这种可变性的发生又向前推进了一步。

在未成熟 B 细胞的分化过程中，IgM 首先表达并暴露在 B 细胞的细胞表面上。一个细胞可以在细胞表面上展示许多拷贝相同的 IgM。每一个细胞可以锚定一个特异的抗原。如果一个特异的细胞遇到其所识别的抗原，其上的 IgM 就会发生多聚化。通过特定的磷酸化信号转导及与 T 细胞的相互作用，细胞的增生便可以被激发，进一步的细胞分化也就开始了。此时大量的 B 细胞就可以分化并形成浆细胞，这些浆细胞将分泌大量的针对某个特异抗原的抗体分子。

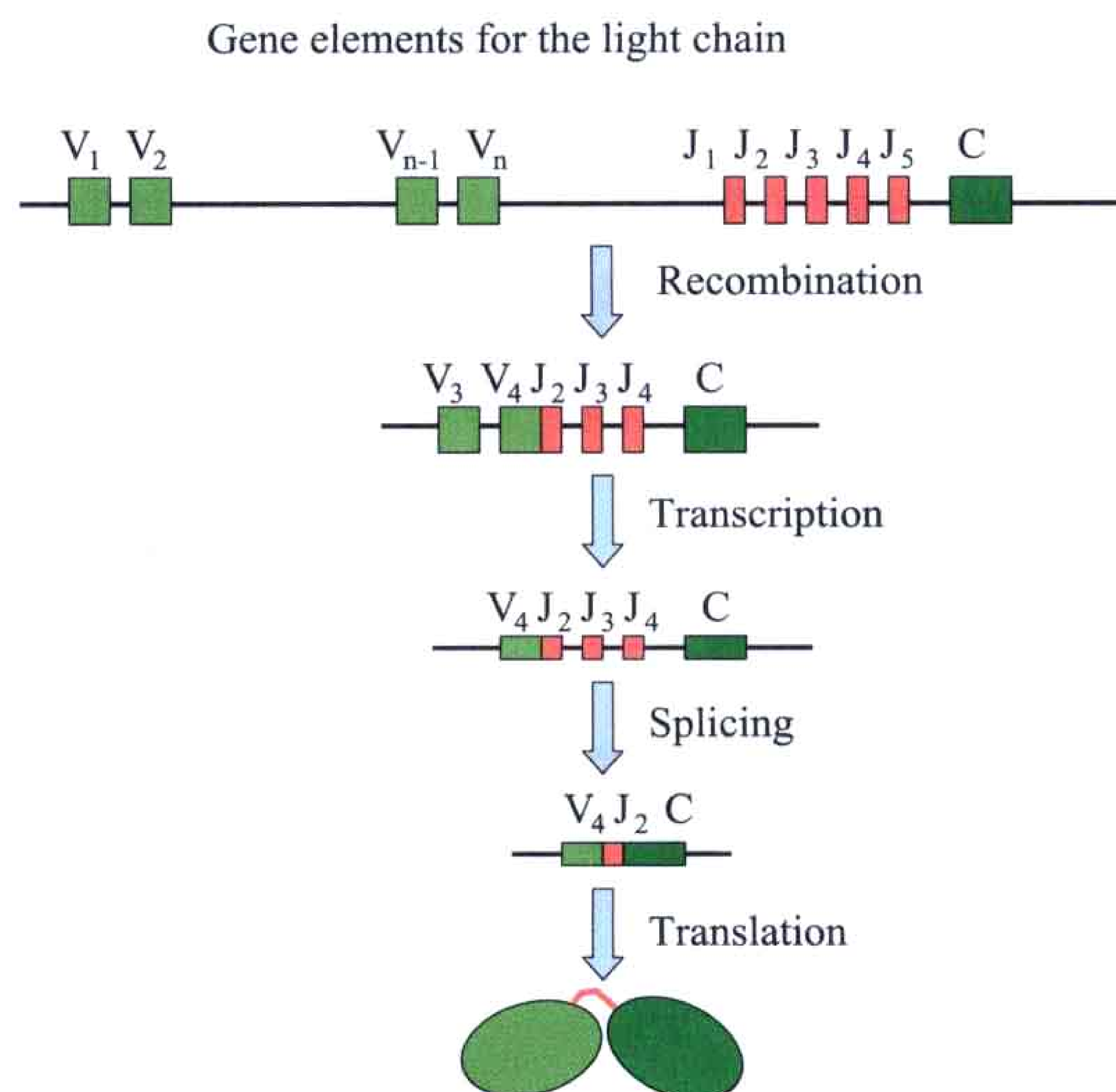


图 14.5　轻链的基因重组和表达。不同的基因片段可以通过很多方式重组起来。未分化的细胞中含有整套的抗体基因。在细胞的分化过程中，一个 V 基因随机地与一个 J 基因连接起来。在随后的 DNA 到前体 mRNA 的转录过程中，某些基因未被转录而进一步被排除。mRNA 被剪切成为成熟的 mRNA 并被翻译成一个特异的轻链。

被某一特定的病原体感染而存活下来的脊椎动物在今后的生存过程中便可以对这一病原体产生免疫。普通的 T 和 B 细胞生命周期较短，但是特异的记忆性 T 和 B 细胞在其所针对的病原体再次入侵时，就可以迅速地产生新的细胞，做出免疫反应。

14.2　细胞免疫——T 细胞介导的免疫系统

T 细胞介导的免疫系统包括抗原呈递细胞和 T 细胞。抗原呈递细胞在其表面上分布有主要组织相容性复合体（major histocompatibility complex，MHC）蛋白。T 细胞含有 T 细胞受体和一系列其他分子，主要包括 CD3、CD4 或 CD8，这些分子又都与信号转导系统相关联。这其中一些重要的中心分子的结构已经知晓，但是分子间的相互作用仍然只是部分得到了理解。

14.2.1　MHC 对抗原的呈递作用

MHC 分子可以分为两类，即 I 类和 II 类。这两类分子都是异质二聚体，都由 4 个结构域组成（图 14.6）。其中两个结构域属于 Ig 类折叠，而另外两个是 αβ 结构域，后两者共同形成了抗原肽的结合位点。

在 I 类 MHC 分子中，肽结合位点位于重链的 α_1 和 α_2 结构域中。第一个测定的结

构非常有趣，除了看到蛋白质结构外，在抗原识别位点还发现了不同肽段的混合物。轻链中的 β_2 微球蛋白亚基主要与重链的 α_3 结构域相结合。在Ⅱ类 MHC 分子中，由两条链同时参与肽段结合位点的形成（图 14.7）。不过，在两类 MHC 分子中肽段结合位点是由相同的结合模式构成的。结合位点的底部是由 8 条链的 β 折叠片层结构构成的，肽段的两侧都有一段螺旋，形成类似“热狗”的折叠模式。

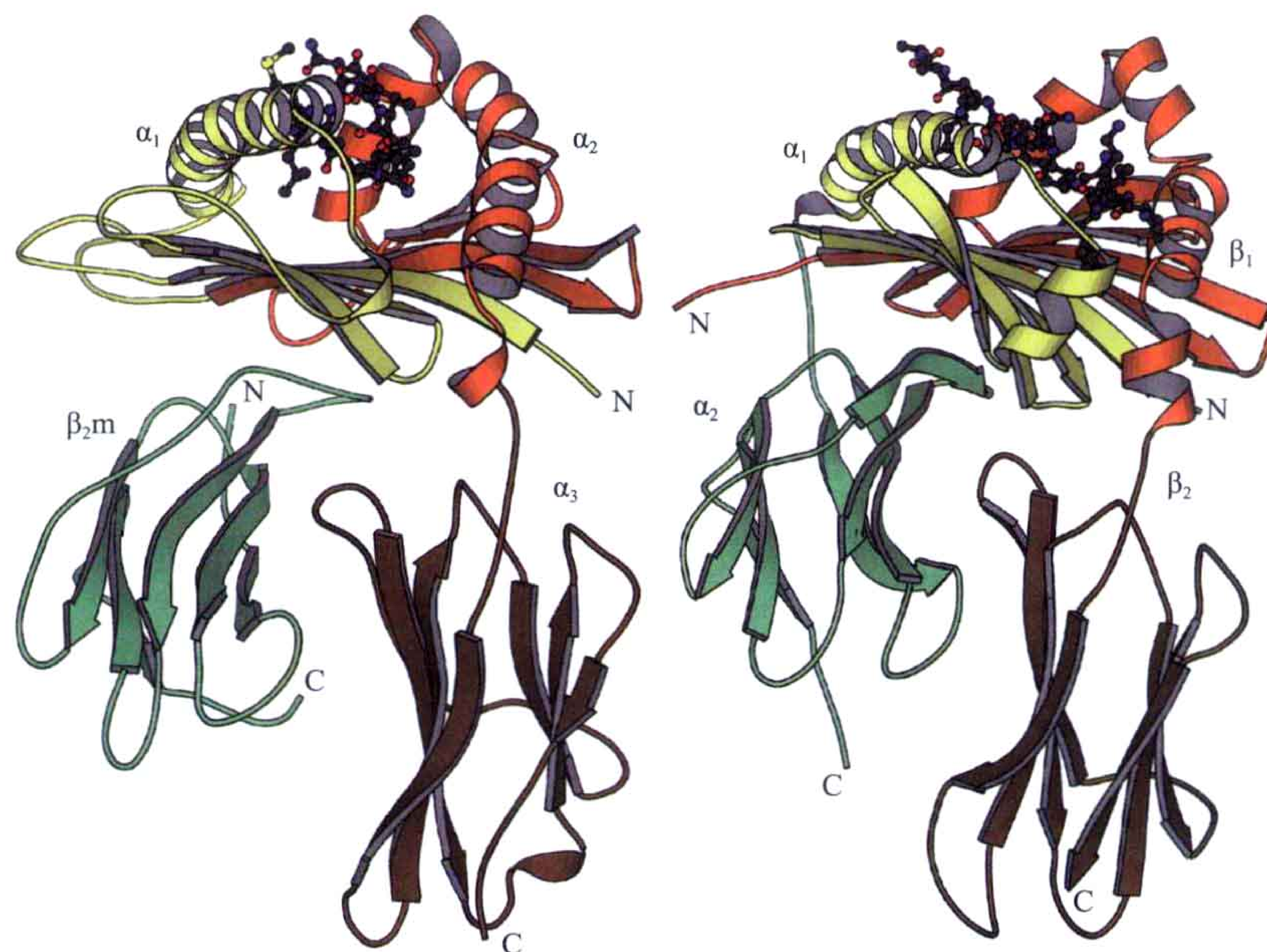


图 14.6　Ⅰ类（左）和Ⅱ类（右）主要组织相容性复合体蛋白结构。Ⅰ类分子含有一个轻链即 β_2 微球蛋白（β_2m，浅绿色）；以及一条含有三个结构域即 α_1（黄色）、α_2（橙色）及 α_3（褐色）的重链，其中前两个结构域（α_1 和 α_2）形成肽结合位点。第三个结构域和 β_2 微球蛋白都属于免疫球蛋白的不变区折叠（PDB: 1A1M）。Ⅱ类 MHC 分子拥有 4 个相似的结合域，但由两条链组成，且结构域之间的连接方式与Ⅰ类 MHC 分子（PDB: 1DLH）不同。

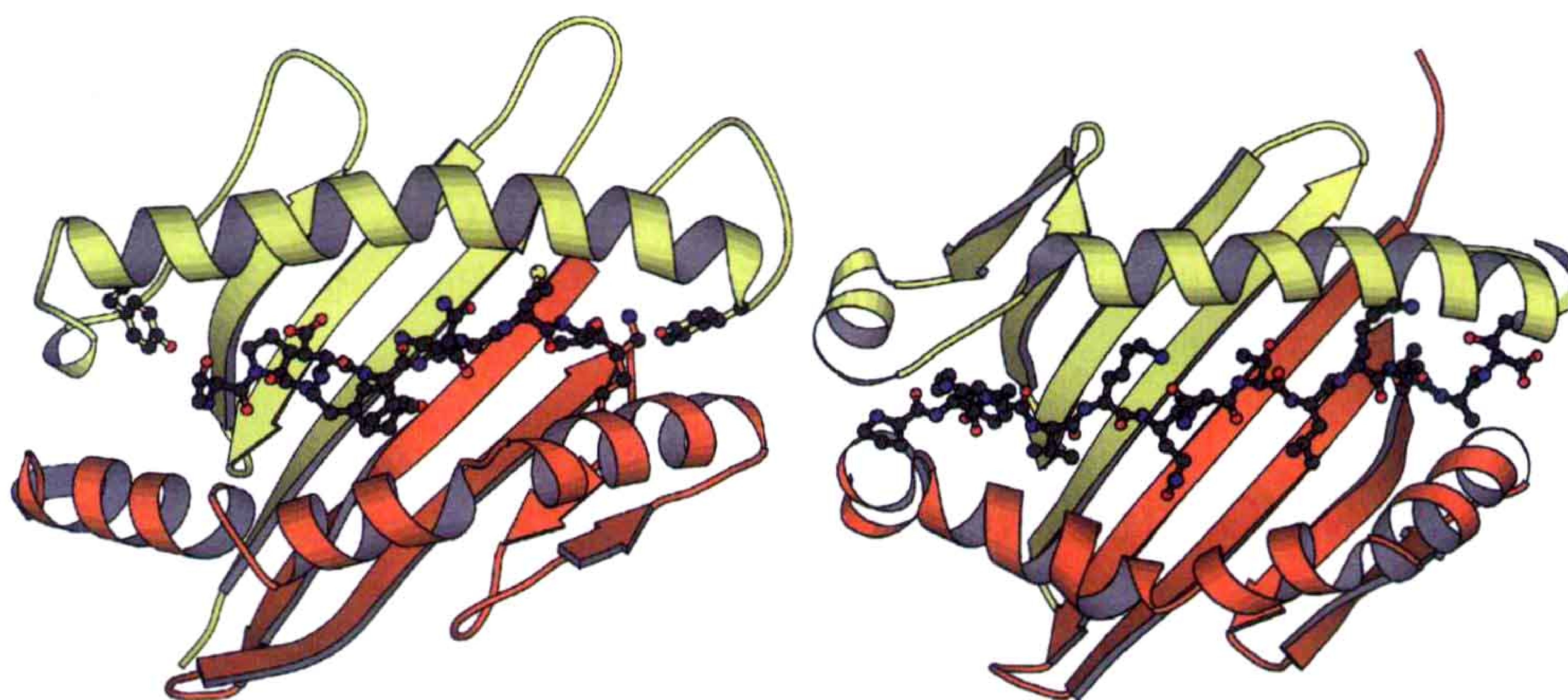


图 14.7　Ⅰ类 MHC 分子（左）和Ⅱ类 MHC 分子（右）与肽段的结合。肽段结合位点是由 8 条链形成的 β 折叠片层和包围在两侧的 α 螺旋所组成的沟槽状结构。结合的肽段用球棍模型显示。在Ⅰ类 MHC 分子中，一些氨基酸封闭了结合沟的两端；而Ⅱ类 MHC 分子中，结合部位的沟槽两端是开放的。

Ⅰ类 MHC 分子呈递在胞质内被降解的抗原蛋白分子的肽段，而Ⅱ类 MHC 分子呈递的肽段来自于细胞外蛋白在内体结构（endosomal compartment）中的降解。

Ⅰ类 MHC 分子通常结合长度在 8 ~ 10 个氨基酸的肽段。肽段的构象比较伸展，锚定氨基酸结合在特定的识别口袋中，这些识别口袋因不同的 MHC 分子等位基因而异。由于结合位点的末端是封闭的，较长的肽段结合时就会引起肽段中部凸起。在Ⅱ类 MHC 分子的结合位点中，结合的肽段形成一个左手多聚脯氨酸螺旋构象。结合位点的两端都是开放的状态，使得较大的肽段可以在位点的两端伸展出去。因此，与Ⅰ类 MHC 分子相比，Ⅱ类 MHC 分子可以结合更长的肽段。

一些非典型的 MHC 分子可以结合糖脂和脂肽并呈递给 T 细胞。不同的 MHC 分子结合位点的多样性满足了需要呈递大量的、不同肽段分子的需求（图 14.8）。与 MHC 分子结合的肽段分子中，一些氨基酸侧链被暴露使得其可以与 T 细胞受体相互作用。

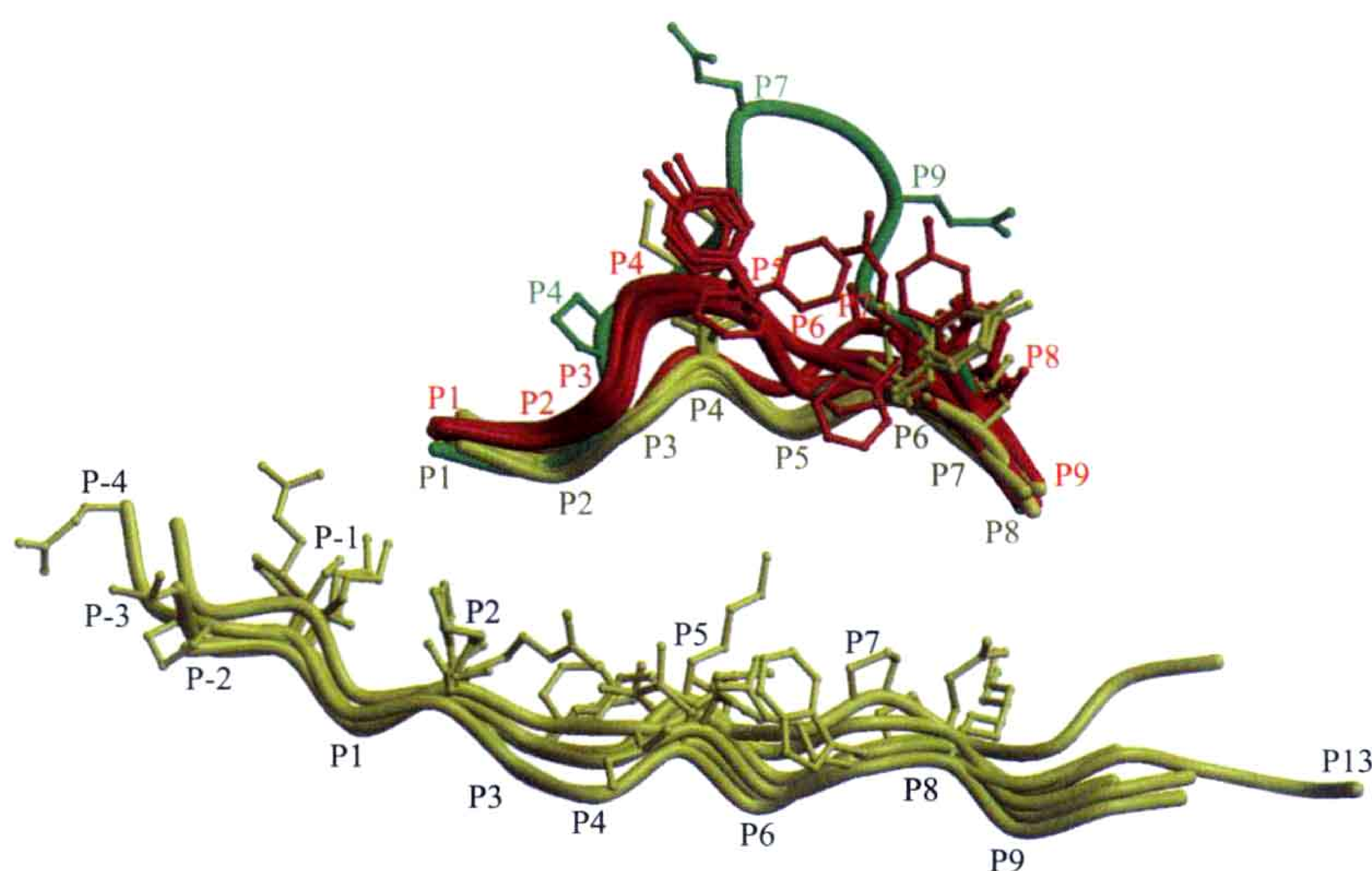


图 14.8　与Ⅰ类 MHC 分子（上）和Ⅱ类 MHC 分子（下）相结合的肽段的结构。不同 MHC 分子的 β 折叠片层被叠合到一块（图中没有显示），它们位于肽段的下方。Ⅰ类 MHC 分子结合的肽段根据长度的不同所示颜色也不同：8 个残基（黄色）、9 个残基（红色）、13 个残基（绿色）。在Ⅰ类 MHC 分子中，结合沟槽的两端是封闭的，因此，大于 8 个残基的肽段结合时就会发生凸起 [原图来自 Rudolph MG 等的文章中 (2006) How TCRs bind MHCs, peptides and coreceptors. *Annu Rev Immunol* **24:**419-466.]

14.2.2　T 细胞受体

T 细胞受体（T-cell receptor, TCR）分布于 T 细胞的表面。除了跨膜区域和一段短的胞质尾巴区域，T 细胞受体还含有一个与抗体分子 Fab 片段相同的结构域。T 细胞受体含有一个不变和可变的 Ig 样结构域，由 α 和 β 或 γ 和 δ 链构成，它们都以与 Ig 分子相似的模式由二硫键连接。αβTCR 与和 MHC 分子相结合的抗原性的肽段相互作用，γδTCR 与病原体介导的糖蛋白或非典型 MHC 分子直接相互作用。与抗体的 Fab 命名相似，TCR 上与 MHC 分子及其所结合的肽段相结合的区域被称为互补决定区（CDR）。

互补决定区与被结合的肽段上的暴露残基侧链相结合，但也与包埋该肽段的 MHC 分子 α 螺旋相互作用（图 14.9）。α 链的可变结构域（V α）与抗原性肽段的氨基端部分相互作用，而 V β 与抗原性肽段的羧基端部分相互作用。这样的结合模式使得 TCR 结合方位有可变性，但抗原多肽相对于 T 细胞受体位于对角线方向。T 细胞受体与 MHC 分子的相对取向对于 T 细胞信号传递可能起重要作用，但是对于信号是如何传递到细胞内部的目前尚不清楚。

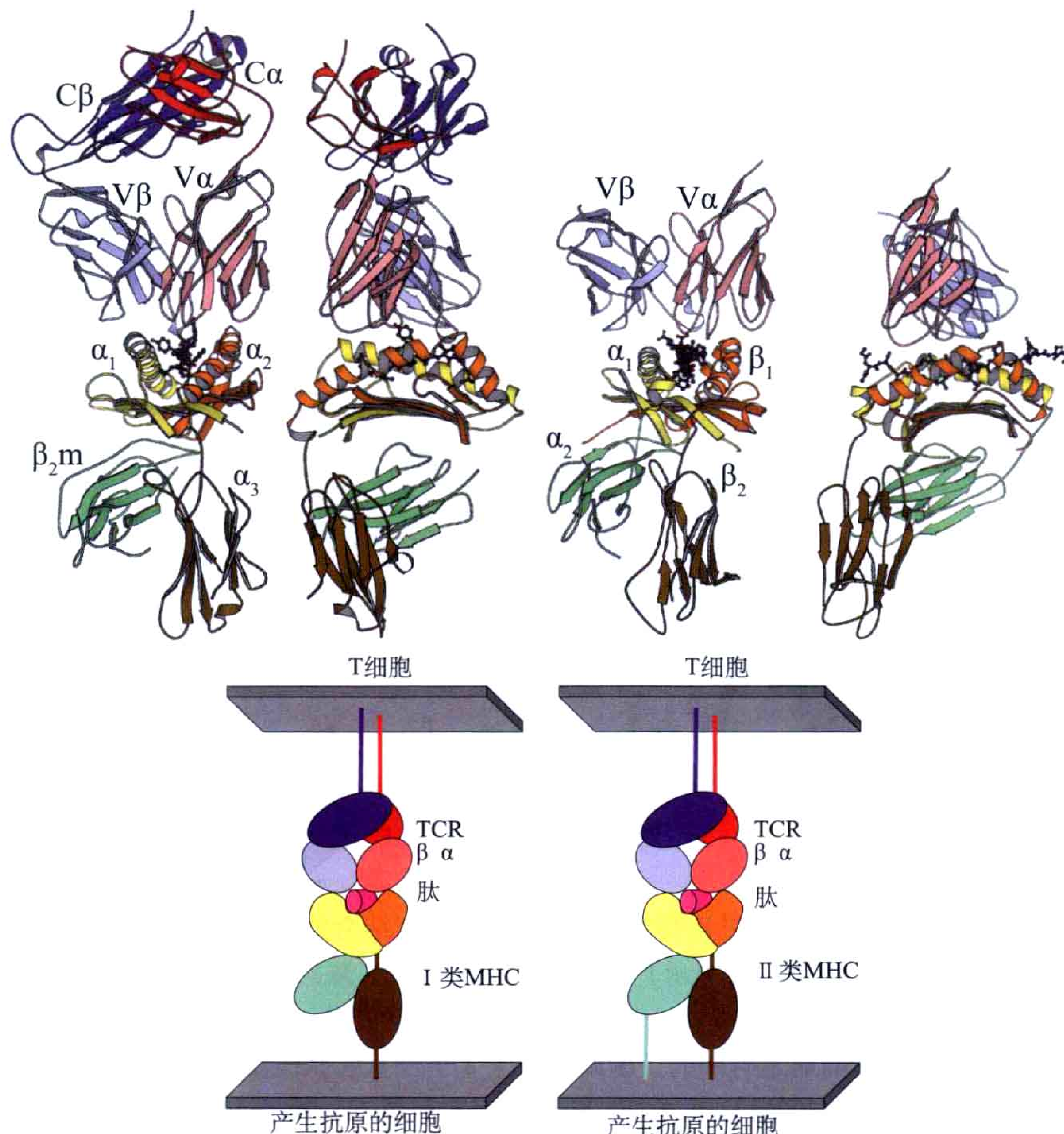


图 14.9 MHC 分子与 T 细胞受体间的相互作用。左：结合在 HLA-A201（Ⅰ类 MHC）上的 T 细胞受体 α 和 β 链胞外的 V 和 C 结构域的两个相垂直的视角。一条病毒肽段结合到 MHC 分子上。右：T 细胞受体的可变结构域 D10 的 α 和 β 链结合于 MHC I-Ak（Ⅱ类 MHC 分子）。MHC 分子在图中位于下方，可以看到结合的肽段位于结合的沟槽中（PDB: 1BD2 和 1D9K）。分子的结合示意图示位于下方。

14.2.2.1 CD8 或 CD4 辅助 TCR 与 MHC 的相互作用

T 细胞受体与 MHC 的相互作用受到辅助受体（co-receptor）CD4 和 CD8 的促进，TCR、CD4 和 CD8 都锚定在 T 细胞的细胞膜上。大量的晶体结构描述了这些分子，以

及这些分子间的相互作用。CD8 是一个异质二聚体（含有 α 和 β 亚基），每一个单体结构含有一个 Ig 结构域、一段长的链接区和一个跨膜螺旋。CD4 是一个单体蛋白，包含 4 个 Ig（D1 ~ D4）结构域，其中 D1 与 Ⅱ 类 MHC 相互作用。

CD4 和 CD8 分别与 Ⅱ 类 MHC 和 Ⅰ 类 MHC 分子底部的几乎相同的保守区域相结合，该区域位于 MHC 肽段结合位点的相反部分（图 14.10）。CD4 也是 HIV1 入侵细胞的主要作用对象。CD4 与病毒刺突蛋白 gp120 相互作用。CD4 与 gp120 相互作用的界面恰是其与 Ⅱ 类 MHC 相互作用的界面，但是该部分与 gp120 的相互作用力要远远大于其与 Ⅱ 类 MHC 的相互作用力。

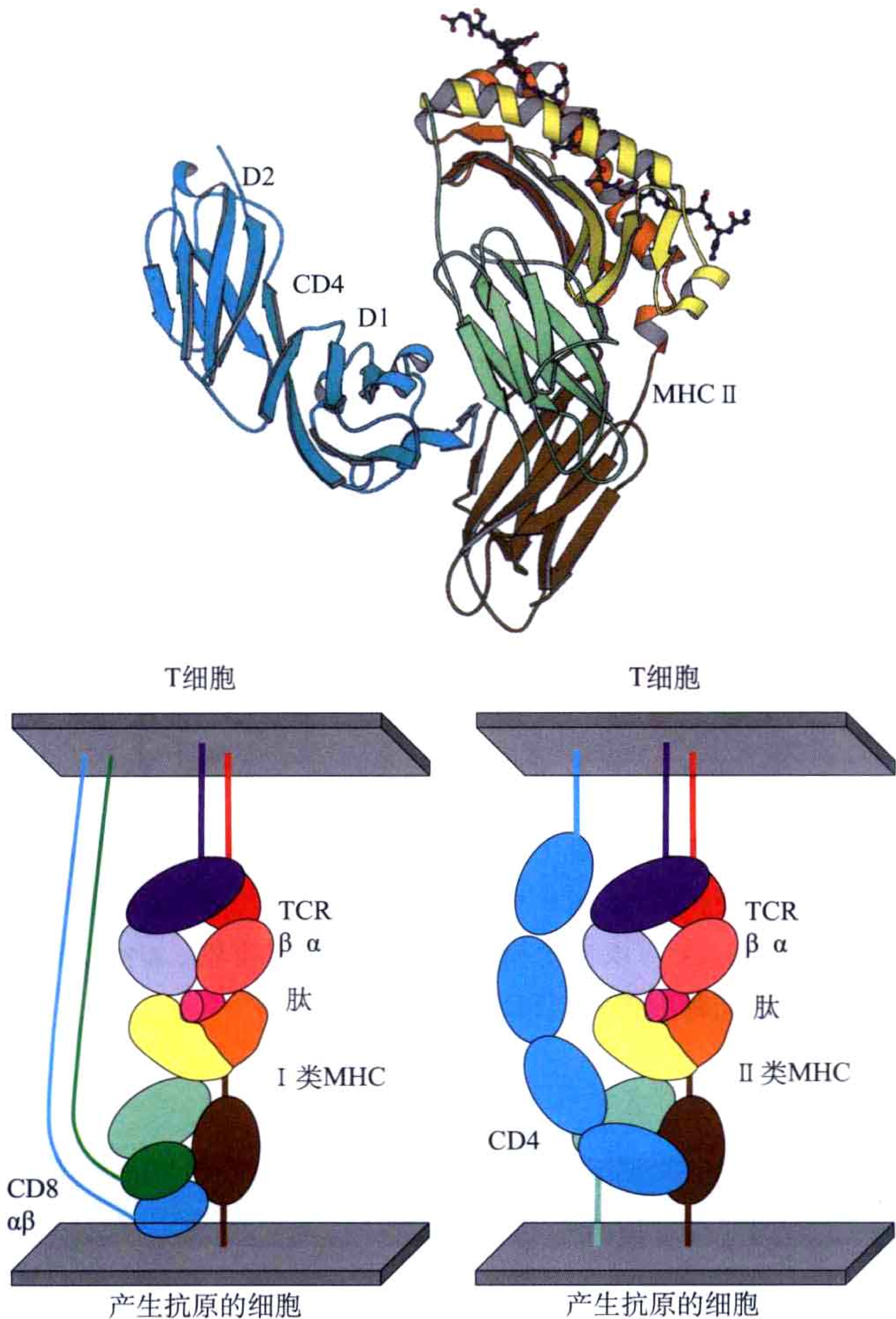


图 14.10　CD4（青绿色）的 D1 结构域与 Ⅱ 类 MHC 分子相互作用。被结合的肽段用球棍模型显示（PDB: 1JL4）。下：TCR 的辅助受体 CD8 或 CD4 分别与 Ⅰ 类 MHC 和 Ⅱ 类 MHC 相互作用的示意图。辅助受体与 MHC 分子的底部，即与肽段结合位点相反的部位相互作用。

14.2.2.2　CD3 辅助分子传递 TCR 分子状态的信号

TCR 分子的细胞内结构域非常小，不足以将信号传递到细胞中。然而，TCR 与三

类 CD3 辅助分子相结合，CD3 辅助分子拥有参与信号传递的结构域。CD3 由两种异质二聚体（γε 和 δε）构成，该二聚体与 TCR 的两条链及另一类同质二聚体（ζζ）相互作用形成一个由 8 条链组成的复合物，每条链都穿越细胞膜。CD3γε 和 CD3δε 异质二聚体的胞外结构域包含并排相互作用的 Ig 折叠（图 14.11）。ζζ 的胞外部分仅含有 9 个残基。连接异质二聚体 CD3 分子与跨膜区的肽段相比较短，并且是在半胱氨酸的帮助下形成二聚体。跨膜区域含有保守的带电荷的残基，这些残基对于 TCR 与 CD3 间的相互作用十分重要。CD3 的胞内部分含有短的序列模体，其被称为 ITAM（immunoreceptor tyrosine-based activation motif，免疫受体酪氨酸活化模体），该模体参与胞内信号的转导。其中所含有的酪氨酰残基可以被 Src 家族激酶所磷酸化。通过受体的聚集及与下游的信号体系的通讯，抗原多肽与 MHC 结合并被 TCR 所识别的信息从 TCR 传递到 CD3 二聚体上。然而，对于胞外与胞内的通讯我们仅了解其中的一部分。从结构的角度上我们依然不能充分地描述细胞是如何识别异己并抗击侵略者的。后续需要更多的工作来阐释在面临重复感染时，T 细胞是如何被介导产生的。

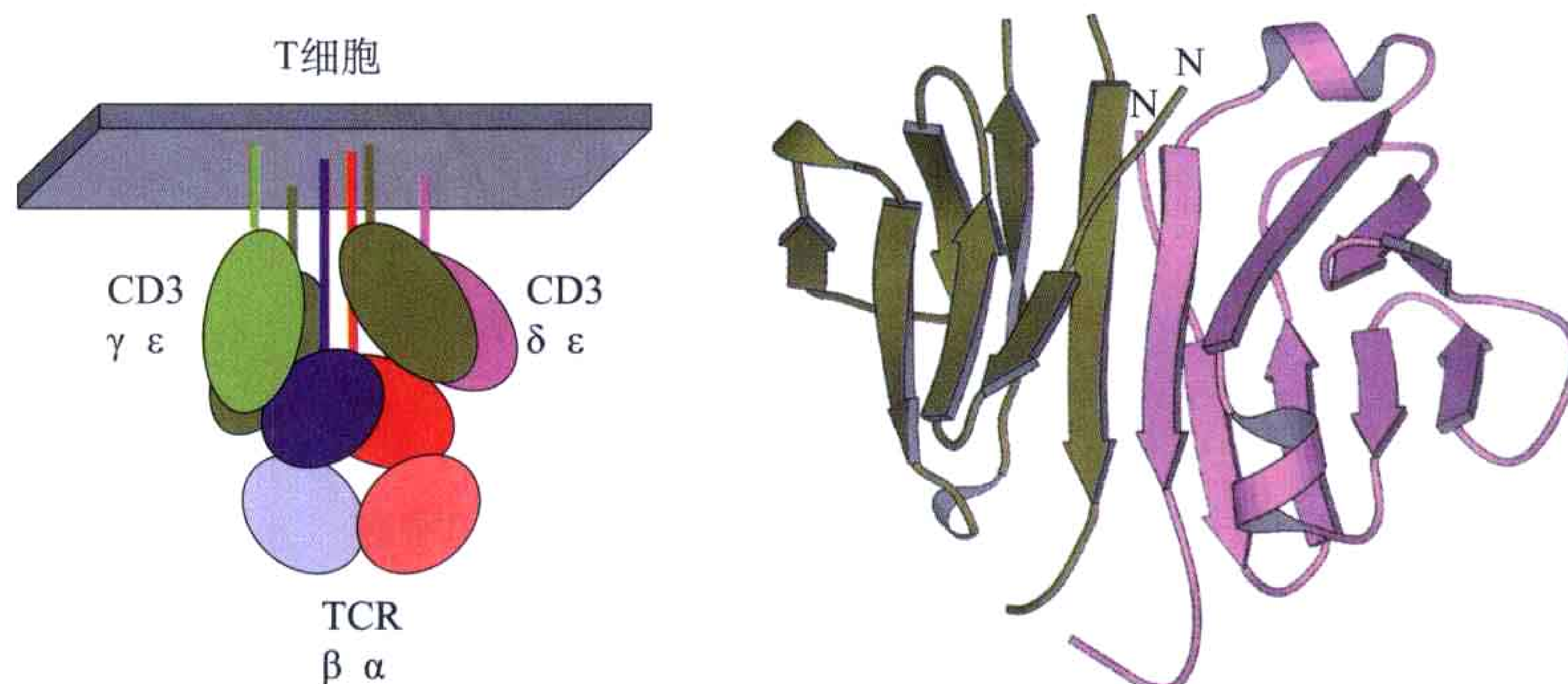


图 14.11　左：T 细胞中 TCR 与 CD3 间相互作用的示意图。CD3γε 和 CD3δε 是异质二聚体并与 TCR 相互作用。它们在膜上的定位使得其与 TCR 可以相互作用，并传递胞内的信号。右：CD3-ε/δ 二聚体胞外结构域的排列方式中含有一个与该视角相垂直的近似的二重轴。ε/γ 二聚体通过相同的方式组合而成。该结构域中的氨基端指向跨膜部分（PDB: 1XIW）。

延伸阅读

原始文献

Amit AG, Mariuzza RA, Phillips SE, Poljak RJ. (1986) Three-dimensional structure of an antigen-antibody complex at 2.8 resolution. *Science* **233**: 747-753.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. (1987). Structure of the class I histocompatibility antigen HLA-A2. *Nature* **329**: 506-512.

Garcia, K. C., Degano, M., Stanfield, R. L., Brunmark, A., Jackson, M. R., Peterson, P. A., Teyton, L., and Wilson, I. A. (1996). An αβ T cell receptor structure at 2.5Å and its orientation in the TCR-MHC complex. *Science* **274**(5285), 209-219.

Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L., and Wiley, D. C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**, 215-221.

综述文章

Rudolph, M.G., Stanfield, R.L., and Wilson, I.A. (2006). How TCRs bind MHCs, peptides and coreceptors. *Annu. Rev. Immunol.* **24**:419-466.

（范雪新 译　王家槐 校）

第十五章
病毒的结构与功能

15.1 病毒的组成

病毒（virus）是一种遗传物质由保护性的蛋白质外壳所包裹着的生命体。有些病毒具有更为复杂的结构，这些病毒的核酸被包裹在一层膜中（enveloped virus）。病毒依赖于宿主进行新蛋白质的合成。通常来说，一类病毒只侵染一类宿主，但也有能够感染包括细菌、古细菌、真菌、植物和动物等各类活细胞的病毒。病毒具有不同的大小和形状（表 15.1）。病毒的遗传物质既可以是 DNA 又可以是 RNA。病毒的核酸可以是单链或双链，它们的基因组可分成一个或几个片段，可以编码多至数百种，少至 4、5 种蛋白质。病毒的结构蛋白组成能够感染宿主细胞的病毒粒子（virus particle），非结构蛋白则负责病毒粒子的有效组装。例如，以 RNA 组成基因组的病毒都能编码一种催化自身遗传物质复制的酶，因为宿主细胞通常不会含有具有类似功能的酶。

表 15.1 具有不同大小和基因组类型的几种常见人类病毒组成

名称	基因组类型[a]	估测的基因组编码的蛋白质数	衣壳组成
脊髓灰质炎病毒(poliovirus)	1 条 ss +RNA 链	4 种结构蛋白，4 种非结构蛋白	二十面体对称，三种蛋白质共 60 个拷贝
流感病毒（influenza virus）	8 条 ss -RNA 链	5 种结构蛋白，5 种非结构蛋白	被膜，外表面由红细胞凝集素和唾液酸苷酶组成
轮状病毒（rotavirus）	11 个 ds RNA 分子	6 种结构蛋白，6 种非结构蛋白	三层蛋白质组成，两内层分别含有 120 和 780 个蛋白质分子
艾滋病病毒（HIV）	1 条 ss +RNA 链	5 种结构蛋白，10 种非结构蛋白	被膜，外表面由糖蛋白 SU 和 TM 组成
天花病毒（variola）	1 个线性 ds DNA 分子	大约 200 种蛋白质	被膜，非常大

a. ss：单链；ds：双链；+ 正义链；－：反义链。也有含有 ssDNA 的人类病毒（细小 DNA 病毒，parvovirus）。

15.1.1 蛋白质衣壳的对称性

包裹着基因组的保护性衣壳（protective shell）是由蛋白质分子组成的（图 15.1）。此外，被膜病毒在核酸和蛋白质衣壳外还有脂双层膜包裹。脂双层膜是在病毒通过出芽（budding）离开宿主时包裹上的，膜上还带有病毒编码的参与组成病毒粒子

外表面的蛋白质。病毒进入宿主细胞的机制就取决于这一件“保护性外衣”的类型：被膜病毒可以通过将病毒粒子自身的外膜融合到宿主细胞膜侵染宿主，但非被膜病毒就只能靠其他机制进入宿主细胞了。

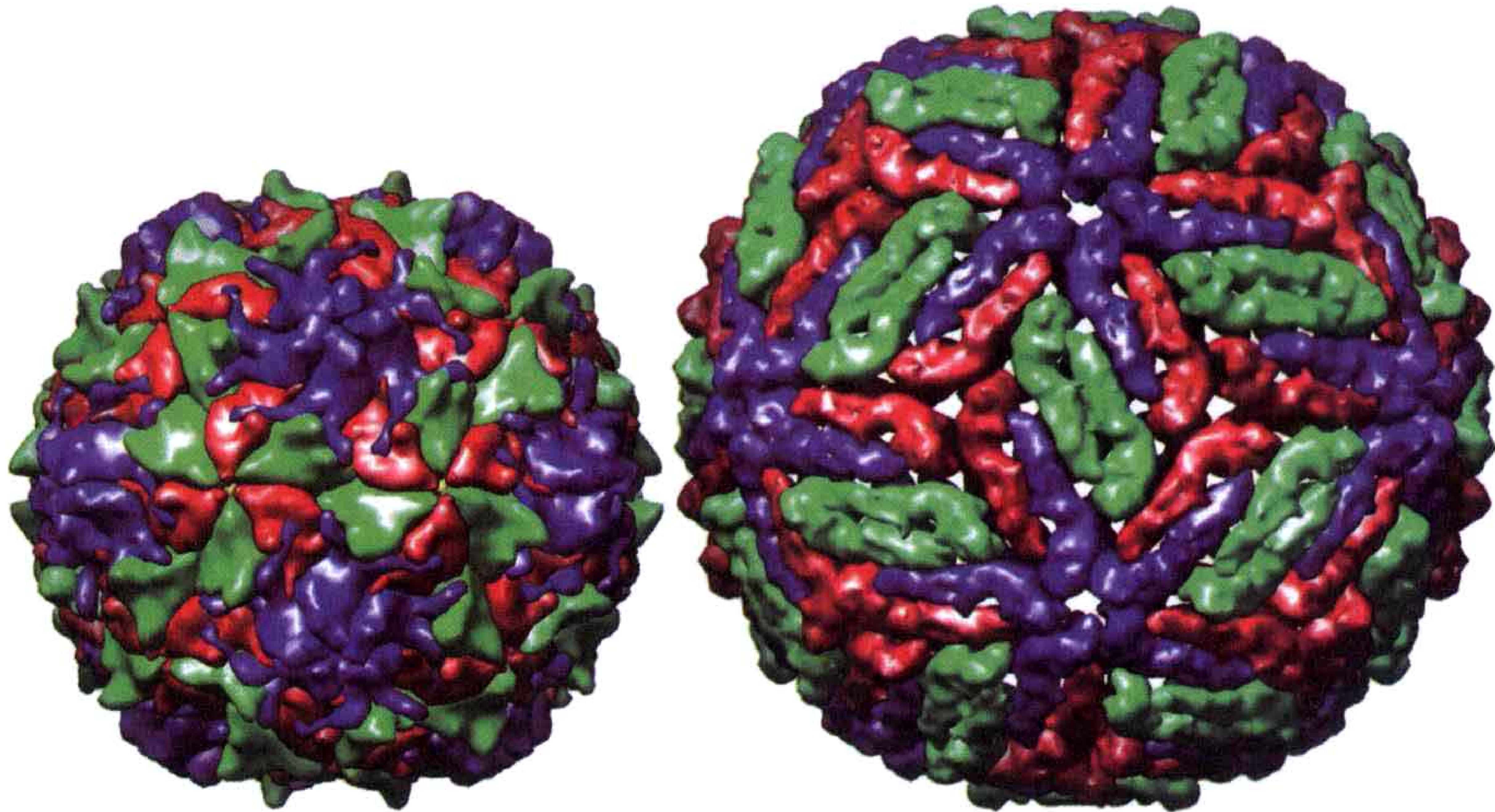


图 15.1　病毒粒子示意图。左：脊髓灰质炎病毒（poliovirus），一种简单的正二十面体病毒，直径约 300Å（基于晶体结构）；右：黄病毒（flavivirus），一种被膜病毒，晶体直接约 470Å（基于冷冻电镜模型与衣壳蛋白晶体结构）。不同颜色代表了不同的蛋白质亚基。来源：VIPER（http://viperdb.scripps.edu/）。

一段核酸分子无法编码一个足够将这段核酸分子包裹住的单独蛋白，因此病毒的蛋白质衣壳是由多个相同的蛋白质分子组装而成的。要保护好病毒基因组，需要一个非常稳定的衣壳。为做到这一点，病毒衣壳的蛋白质分子组装是对称排布的，不同亚基间的相互作用依赖于相同的蛋白质 - 蛋白质相互作用表面。在病毒中发现了两种本质上不同的对称性——螺旋对称和正二十面体对称。螺旋对称导致病毒粒子包装成杆状，如烟草花叶病毒（tobacco mosaic virus）；二十面体对称导致病毒粒子包装成密闭的近似球形的结构。

病毒粒子的形状依赖于衣壳的组成。一些被膜病毒由于外膜的柔性，形状可以发生变化；而其他病毒由于膜蛋白之间或内部的对称蛋白层之间的重复相互作用而具有固定的形状。非被膜病毒通常或为螺旋，或有又有二十面体对称。也有的病毒有更为复杂的形状，如 T4 噬菌体（phage T4）具有二十面体对称的头部以及螺旋对称的尾，头和尾上带有组成纤毛的蛋白质复合体和其他对感染过程具有重要作用的蛋白质结构。

15.1.2　准等价

正二十面体是由 20 个等边三角形组成的物体，它具有五重、三重和二重对称性。要产生一个正二十面体需要 60 个完全等同的单位（图 15.2）。在一个封闭的物体中这是可能产生的最高的对称性（由 12 个五边形产生的正十二面体具有相同的对称性）。

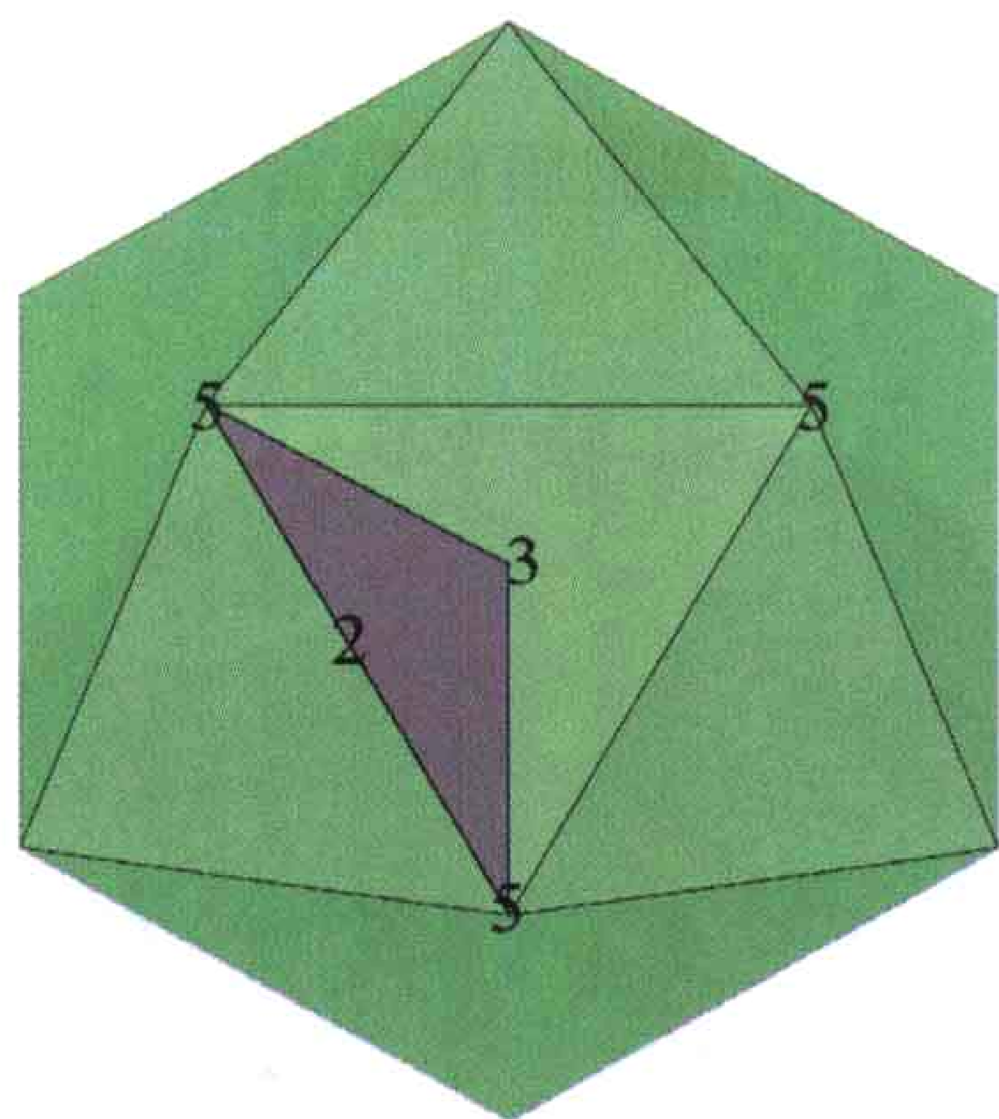


图 15.2　一个正二十面体中的五重、三重和二重对称轴。重复的单位用灰色标记，这种重复的单位还可以有很多种可能的选择。

有的病毒的衣壳是由同一种衣壳蛋白的 60 个分子拷贝组成，但大多数病毒粒子是由大量的相同亚基组成的。在 1962 年详细的病毒结构还不为人知时，Caspar 和 Klug 提出了准等价理论（quasi-equivalence theory），该理论尝试解释了大量的衣壳蛋白分

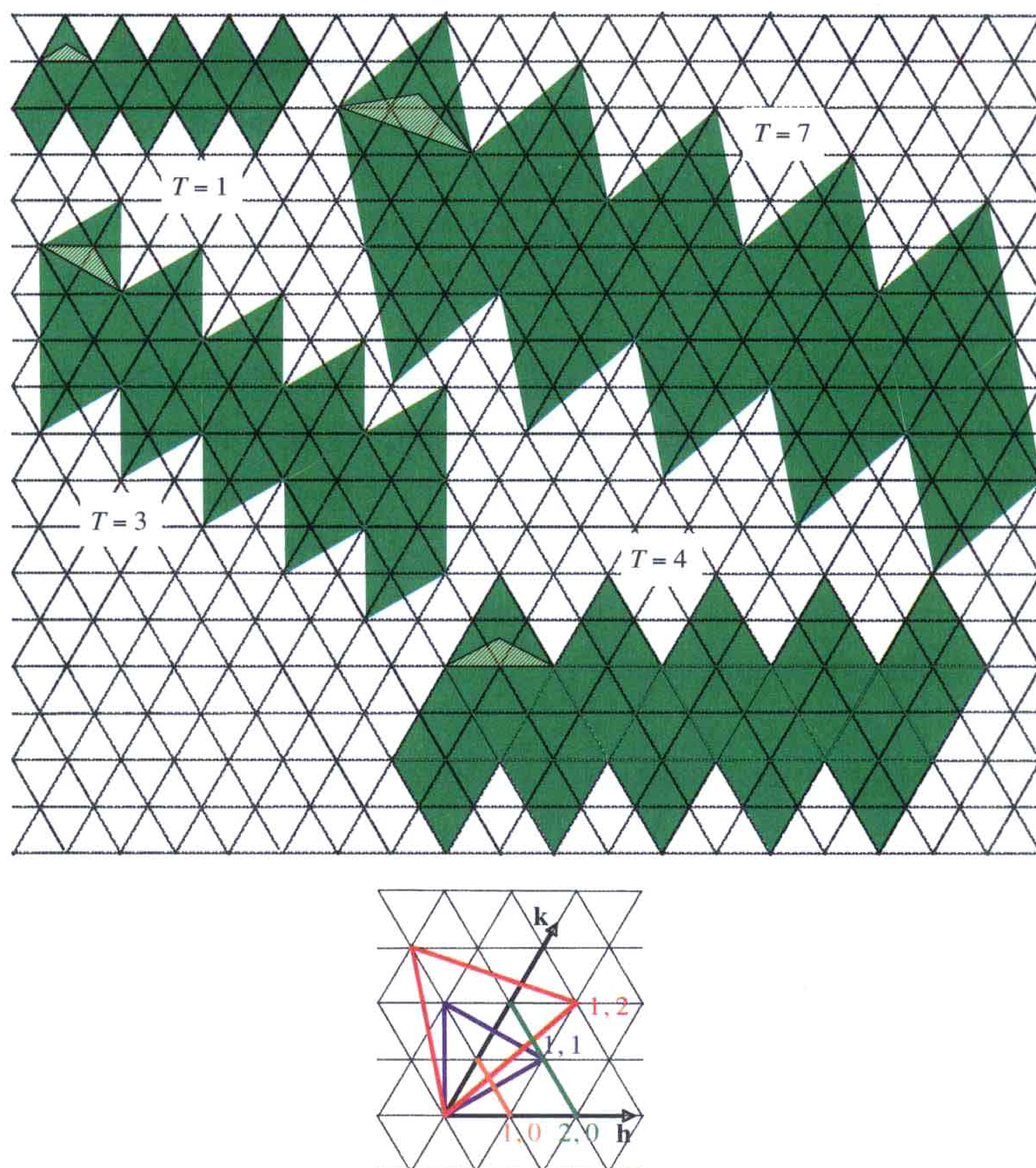


图 15.3　将六重对称换成五重对称的三角网络。20 个三角形（绿色）对应着二十面体的表面，三角形的每一个角对应着一个五重轴。六重对称变成了一个符合正二十面体三重和四重的准六重。在 T=1 时，只有五重对称；T=7 时，准六重轴不符合任何二十面体对称轴。下：h, k 坐标系。对每一个 T 展示了一个三角形（橙色：T=1；蓝色：T=3；绿色：T=4；红色：T=7）。

子是如何组装成正二十面体的。如果一个衣壳含有 60 的倍数个蛋白质亚基，那么相同的蛋白质分子就会处于不同的环境中。该理论基于蛋白质亚基间的接触面是相似的、准等价的，蛋白质亚基的组装使用相同的、只有轻微形变的化学键。该理论的基础是亚基间通过相似的接触面有可能形成六重或五重相互作用。通过一种常规的方式将六重接触面替换成五重接触面，一个平面的三角形网络可以被转变成一个正二十面体。五重接触面产生了曲度，根据这些五重轴位置的不同，可以形成由不同数量三角形组成的正二十面体。Caspar 和 Klug 发现某些 60 倍数个的分子可以以这种方式组装成具有正二十面体对称性的结构。这些倍数按公式 $T=h^2+k^2+hk$（h 和 k 是整数）对应着三角划分数 T=1, 3, 4, 7, 9, 13, 19, 21, 25 等（图 15.3）。衣壳中亚基总数量则是 60T。

准等价理论对蛋白质亚基排布的预测已经大部分被病毒结构研究所证实。T=1, 3, 4, 7 和 13 病毒粒子的晶体结构的解析使得在原子水平分析准等价理论成为可能（图 15.4）。有些情况下，蛋白质亚基接触面确实像 Caspar 和 Klug 预测的一样符合准等价理论。而在另一些情况下，即使病毒表面蛋白质亚基的位置符合理论的预测，准等价接触面也主要是在不同组原子间形成的。也有一些情况与预测的接触面相似性相去甚远。在 T=7 的多瘤病毒（polyomavirus）和 SV40 中，六重位置和五重位置是由亚基的五聚体占据的，病毒粒子因此具有 360 个蛋白质亚基而不是预测的 420 个。在蓝舌病毒（bluetongue virus）这样的含有双层蛋白的病毒中，内层含有 120 个拷贝分子的衣壳蛋白，对应着 T=2，这在准等价理论中是不会出现的。对于这些病毒，两个亚基间的相互作用有着极大的不同。



图 15.4 T=1, 3, 4, 7 和 13 的病毒粒子的相对大小。病毒粒子表面的阴影表示与中心的距离，距离越近越暗。有的粒子呈二十面体，但所有粒子都具有二十面体对称性。图示基于的病毒晶体结构分别为（从左至右）：卫星烟草坏死病毒、噬菌体 MS2、*Nudaurelia capensis* 的 ω 病毒、噬菌体 HK97 和蓝舌病病毒。来源：VIPER（http://viperdb.scripps.edu/）。

15.1.3 粒子组装和稳定性的控制

在很多简单的病毒中，粒子不需要其他分子的协助就可以形成，因此这种自组装一定有某种内在属性。正二十面体病毒的衣壳蛋白分子主要负责病毒粒子的大小和形状，并且能形成五重、三重和二重对称接触面。当 60 的倍数个化学性质上相同的亚基组成衣壳时，分子一定能够以一种正确的方式形成包裹有序的二十面体对称的衣壳，尽管在有些接触面上存在少许差异。第一个解析的病毒粒子的结构是 T=3 的一种植物

病毒。通过其结构能够看到衣壳蛋白由一个球形结构域（globular domain）和一段多肽链 N 端的延伸部分（extended segment）组成。

与大多数其他非被膜二十面体病毒一样，衣壳蛋白的球形结构域是一种由两个反平行的 4 股 β 折叠片形成的果冻卷拓扑结构（jellyroll topology）（图 15.5）。片层的长度及连接片层的 loop 的长度和构象在不同病毒间有着极大的差异。在某些病毒中，果冻卷式的折叠甚至会被插入序列中间的另一个整个结构域所打断。

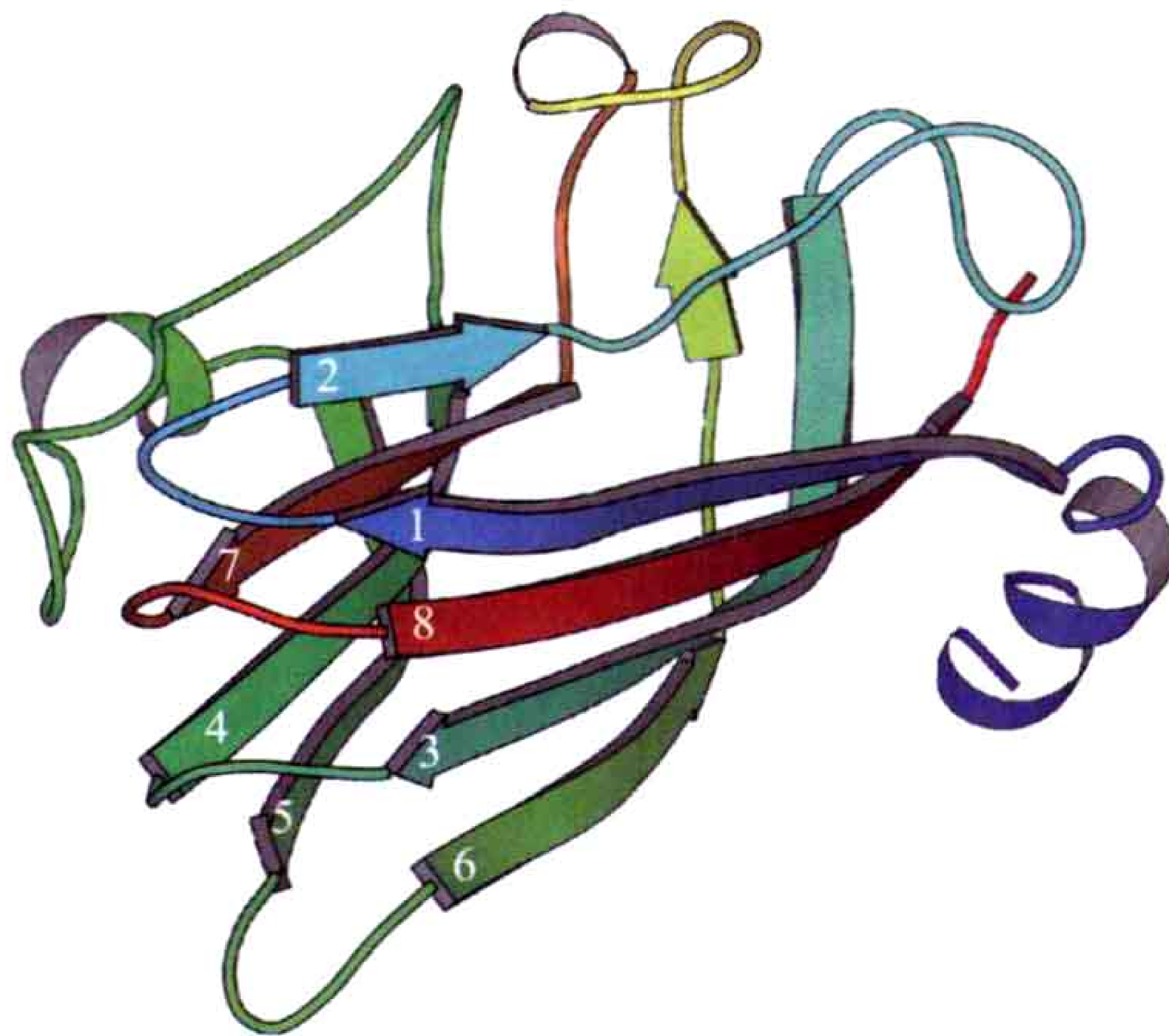


图 15.5 病毒衣壳蛋白亚基中的果冻卷折叠（卫星烟草坏死病毒，PDB：2BUK）。

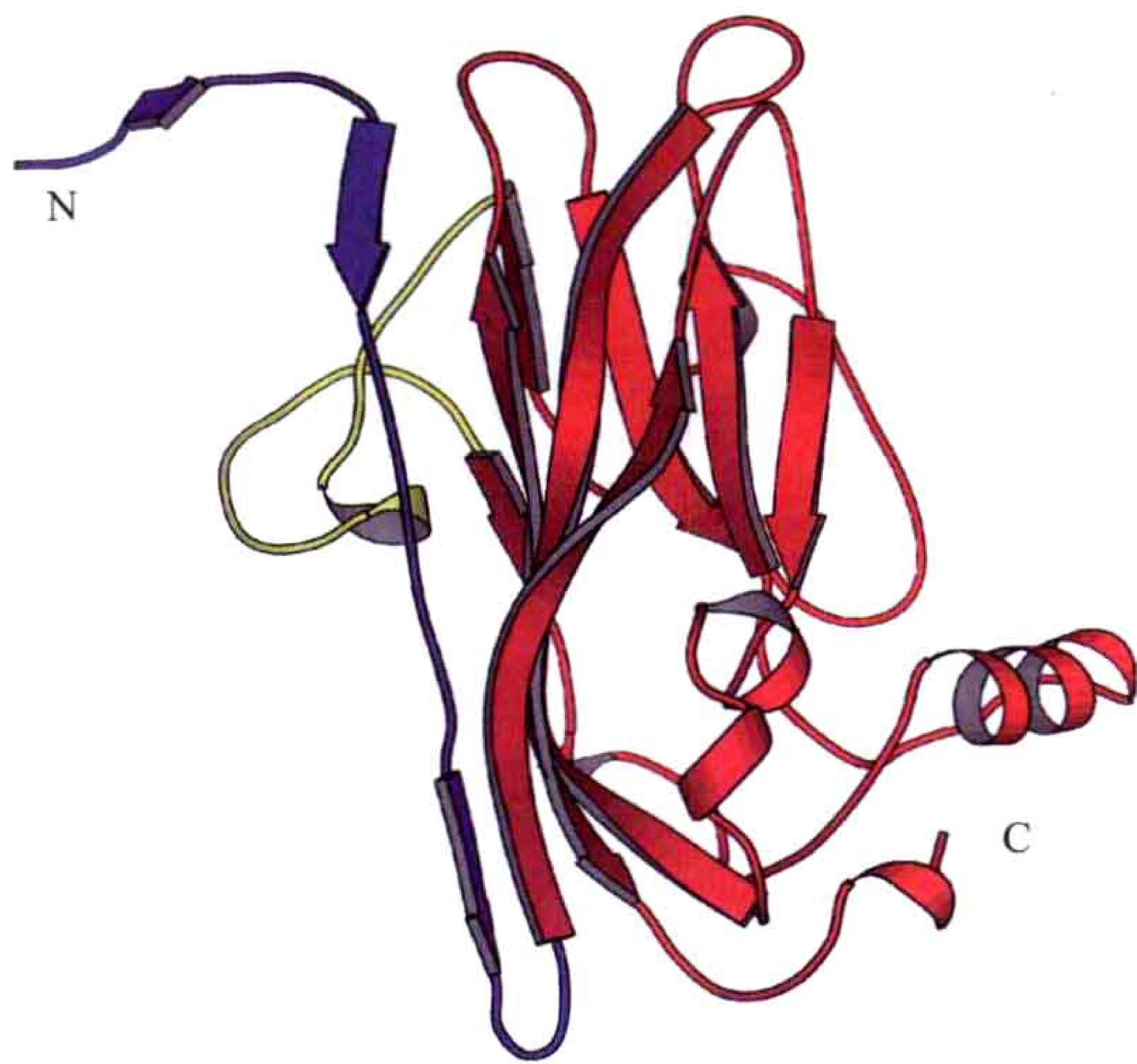


图 15.6 病毒（南部豇豆花叶病毒）衣壳蛋白的果冻卷结构（红色）和 N 端延伸部分（蓝色）。C 端部分为部分有序结构（PDB：4SBV）。

在一种 T=3 的病毒中，完全相同的单个蛋白质亚基会处于三种不同的相互作用中。例如，其中一个亚基形成五重接触面而另两个在正二十面体三重轴的地方形成准六重接触。在第一个被研究的 T=3 的植物病毒中，衣壳蛋白的延伸部分被用来稳定亚基组装时的正确排布。在三个亚基分子中，其中两个的延伸部分是完全无序的，而另一个却是部分有序的（图 15.6）。这些有序部分插入一些亚基的交界面之间，并在三重对称轴的位置发生相互作用形成所谓 β 套环（beta annulus）的结构（图 15.7）。在有序部分插入的交界面处，亚基 - 亚基间的相互作用与无序部分形成的准等价相互作用是完全不同的。N 端的延伸部分就像一个能控制组装使之达到一定曲度的开关。N 端延伸部分这种有序 / 无序的开关机制对病毒粒子的大小十分重要。该部分被敲除的突变体只能形成 T=1 的病毒粒子，而不能形成 T=3 的粒子。

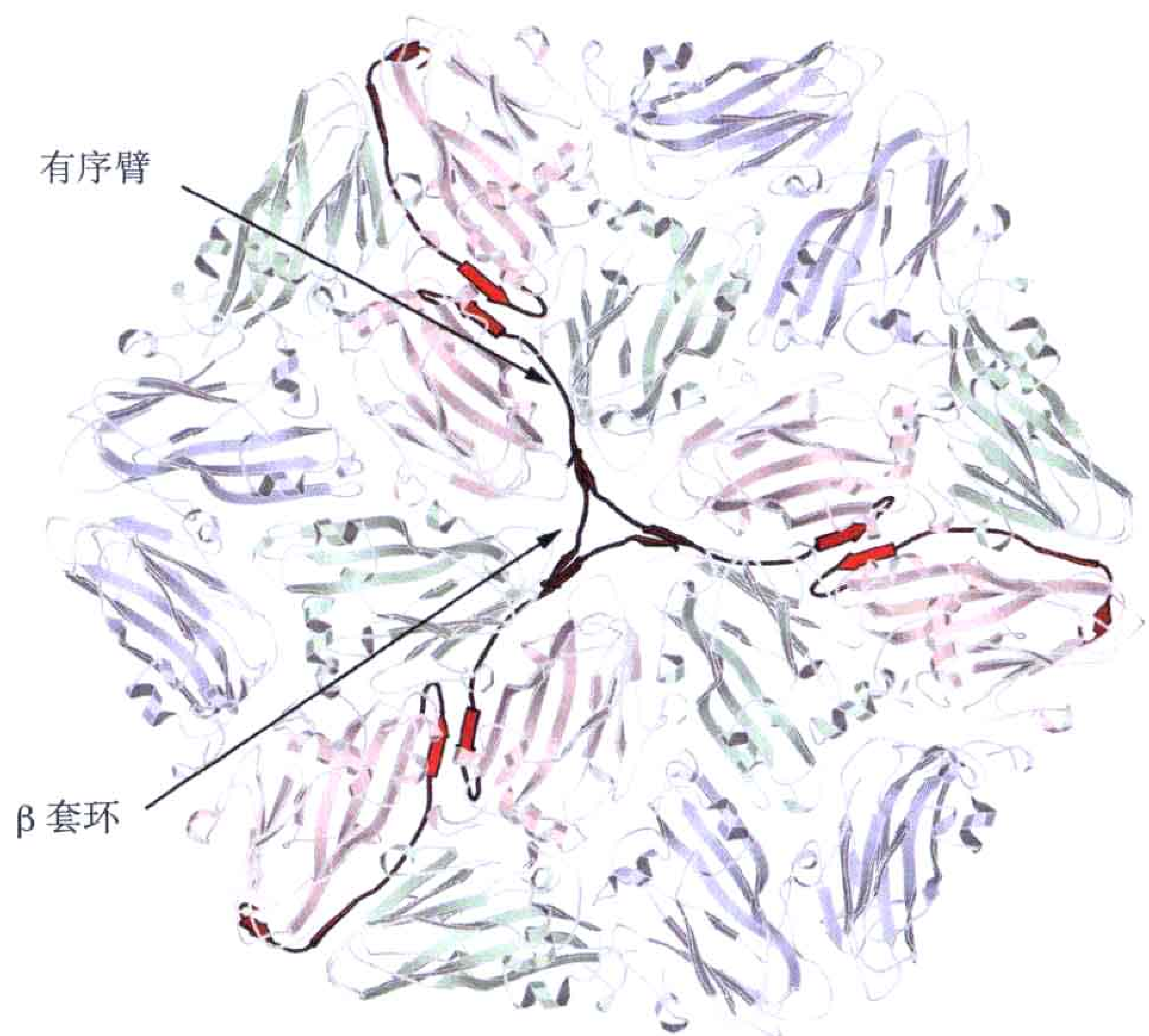


图 15.7　病毒（南部豇豆花叶病毒）中 18 个亚基围绕三重轴（准六重轴）的排布。亚基中部分有序的结构（红色）与对称亚基在三重轴处相互作用形成 β 套环（箭头指示处）。在这种病毒中，所有亚基中的 N 端 23 个残基都是无序的。这一部分的带正电荷的残基很可能与非对称的病毒 RNA 相互作用。

这种延伸部分的有序 / 无序开关机制存在于大多数非被膜二十面体病毒中，包括 T=3，T=4 和 T=7 的病毒粒子。但是这些病毒中开关机制的细节却并不相同。也有具有准等价排布的病毒不含有延伸部分的例子。在这种情况下，衣壳蛋白的折叠不是普通的果冻卷式的，因此延伸部分的存在可能是果冻卷折叠的特点。而无序部分存在如此广泛的原因很可能是不同衣壳蛋白具有共同的祖先，而并非该部分在功能上的不可或缺性。

15.2　病毒进入宿主细胞的机制

15.2.1　膜融合

在进入宿主细胞的方式上，病毒已经进化出了多种机制。一旦进入宿主细胞，它们就会关闭宿主细胞本身细胞组分的产生，并操纵细胞机器产生新的病毒粒子。动物病毒需要穿过宿主细胞的细胞膜，对于被膜病毒，这是通过膜融合（membrane fusion）实现的：病毒通过将自身的膜与细胞膜连接，使含有病毒基因组的核衣壳（nucleocapsid）进入宿主细胞内部（见 4.4.3 节）。根据病毒基因组特点的不同，DNA 病毒会进入宿主细胞的细胞核，而 RNA 病毒则会留在胞质中。

病毒进入宿主细胞的整个过程都是由病毒控制的。病毒粒子本身只是几种分子组成的简单复合物，各种进入机制取决于各种与宿主细胞的相互作用引起的不同的激活

机制，与宿主细胞的相互作用一般会导致病毒粒子构象上的变化。

一些被膜病毒，如人类艾滋病病毒，可以直接与细胞膜融合。而另一种进入宿主细胞的途径是内吞（endocytotic pathway）。病毒通过与宿主细胞表面的受体结合，形成含有完整病毒粒子的内吞颗粒进入细胞。进入细胞的内涵体（endosome）内部的 pH 比较低，这通常引起病毒某些构象上的变化。对于像流感病毒这样的被膜病毒，当病毒的膜与宿主细胞膜融合时就会使核衣壳释放进入宿主细胞。而对于像腺病毒（adenovirus）和呼肠孤病毒（reovirus）这样的非被膜病毒，核衣壳从内涵体释放到宿主细胞内的机制尚不清楚。

被膜病毒通过融合肽段（fusion peptide）将病毒自身的脂双层膜与宿主细胞膜连接。这些肽段通常是病毒表面某些蛋白质的一部分，但平常并不暴露在外而是折叠在蛋白质内部或参与其他蛋白质的相互作用。对于不同的病毒，这些肽段如何暴露出来并结合到宿主细胞表面造成膜融合的机制也各不相同。

融合肽段可以分为两种类型，第一类是在流感病毒中发现的。红细胞凝集素（hemagglutinin）是一个很大的糖基化膜蛋白，它以三聚体的形式在病毒表面形成刺突。形成病毒粒子时，该蛋白质会在特异性位点被剪切，形成一个亚稳态结构。C 端的 HA2 结构域将蛋白质定位在病毒膜上，并形成一长一短两条螺旋结构，两条螺旋结构之间被一条很长的 loop 连接（图 15.8）。部分长螺旋结构与三聚体中其他亚基的相同部分相互作用形成螺旋卷曲结构。在 HA2 的 N 端就是融合肽段。N 端的 HA1 形成受体结构域并沿着 HA2 延伸。

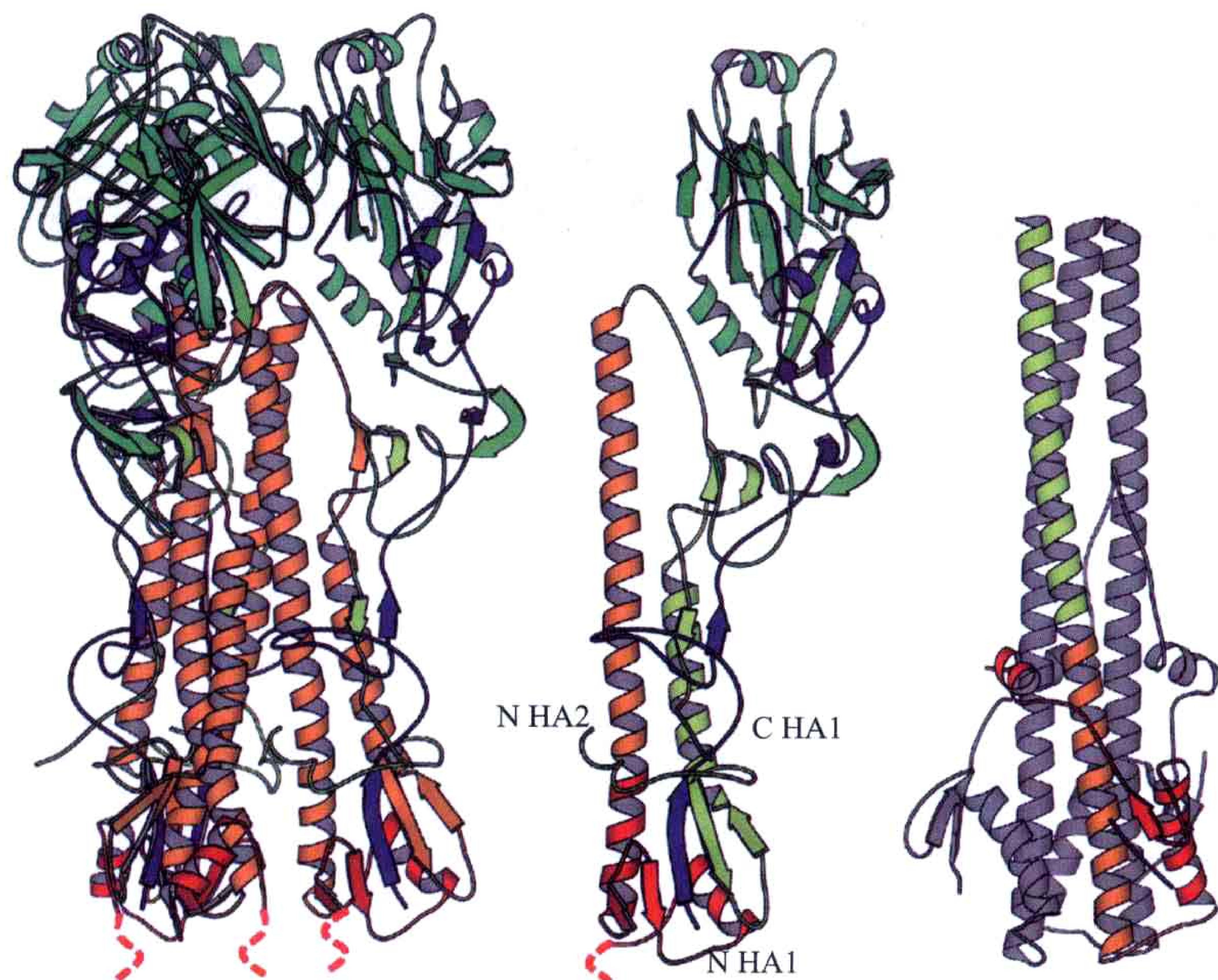


图 15.8 流感病毒红细胞凝集素。左：在中性 pH 下剪切成 HA1 和 HA2 肽段后的三聚体（PDB：1RU7）。膜结合在 HA2 的 C 端（在末端红虚线处）。中：中性 pH 下的单体结构。融合肽段（N HA2）位于分子中部，并与其他亚基相互作用。右：低 pH 下的蛋白质结构（PDB：1HTM）。融合肽段变为无形结构，但必须定位在三个螺旋结构的顶部。在低 pH 条件下长螺旋的构象发生了很大的变化，下部结构（红色）发生弯曲并包裹到长螺旋结构的侧面；上部结构（黄绿色）形成一条长的环形结构和一条螺旋结构。

被内涵体内的低 pH 诱导后，红细胞凝集素的构象发生了非常大的变化。在中性 pH 时呈环状结构的部分延长了分子中的长螺旋区域，导致了融合肽段的暴露。在另一端，一部分螺旋结构发生弯曲并与长螺旋呈反平行走势，而这段区域正是与膜相连的部分。这种对折的构象变化被认为能够使病毒的膜靠近宿主细胞膜从而使融合肽段插入。

另一类融合肽段是在黄病毒（flavivirus）和甲病毒（alphavirus）中发现的，它们都是有着正二十面体对称性的被膜病毒。这类病毒的融合肽段平常也是通过蛋白质间的相互作用被包裹在膜内部。它们负责进入宿主细胞的膜蛋白分别称为 E（黄病毒）和 E1（甲病毒属）。这两个蛋白质在构象上与流感病毒中的红细胞凝集素完全不同（图 15.9）。这两个蛋白质中的融合肽段是两个 β 折叠束之间的一段环状结构。在未被激活的状态下，两个蛋白质以相切的走向形成一个二聚体，一个亚基中的融合肽段与另一个亚基的一个结构域相互作用。在低 pH 条件下，这种相互作用会发生巨大改变并导致融合肽段暴露出来。这种构象变化的机制尚不清楚，但可能与其他使用融合肽段的被膜病毒中的融合机制类似。

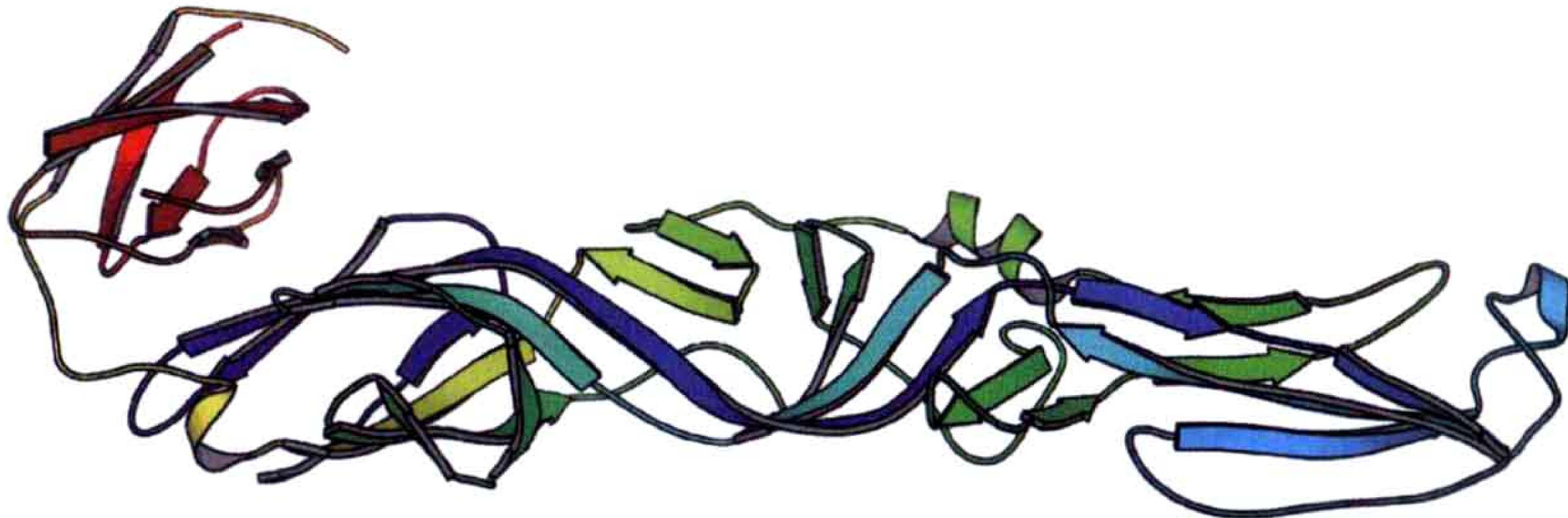


图 15.9　甲病毒属塞姆利基森林病毒的 E1 蛋白结构（PDB：1L9W）。N 端用蓝色表示，C 端用红色表示。融合肽段位于分子最右端并与另一个亚基相互作用形成同源二聚体。C 端是膜定位结构，但在该晶体结构中不包含这段区域。

15.2.2　非被膜病毒进入宿主细胞的机制

非被膜病毒无法利用膜融合进入宿主细胞，而是使用其他机制。一种情况是病毒能够在宿主细胞上形成一个小孔而使病毒粒子或病毒基因组通过。

小 RNA 病毒（picornavirus），包括脊髓灰质炎病毒（poliovirus）和鼻病毒（rhinovirus，普通感冒病毒），是只含有几个结构蛋白和非结构蛋白的简单病毒。它们进入宿主细胞的机制被作为寻找抗病毒感染药物的手段而得以广泛研究。成熟的脊髓灰质炎病毒粒子包括三种衣壳蛋白，分别是 VP1、VP2 和 VP3，在成熟的病毒粒子中这三种蛋白质各含有 6 个拷贝的分子数。这三种蛋白质均含有 jellyroll 折叠和能在粒子内部形成网络结构的 N 端延伸区域（图 15.10）。在蛋白质衣壳内部还有一个单独的多肽称为 VP4，它起初是 VP2 的 N 端延伸的一部分，但在病毒粒子形成后被自催化切割。这种“成熟切割”（maturation cleavage）机制在病毒中广泛存在，病毒粒子为酶切反应的发生提供了良好的环境。虽然酶本身自催化的效率并不高，但对病毒来说并不需要很高的催化周转率。

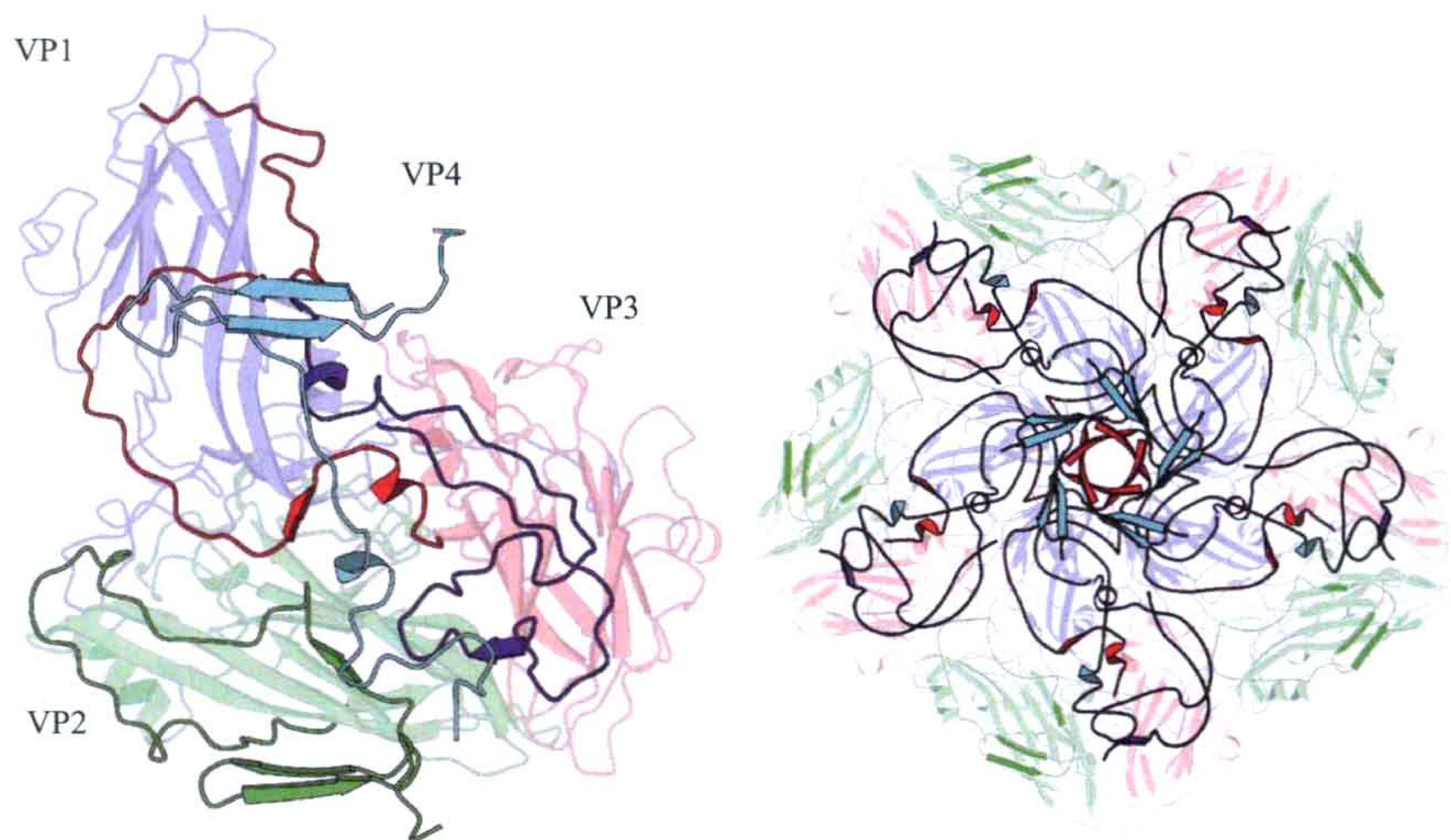


图 15.10 脊髓灰质炎病毒衣壳结构的 N 端部分。左：从细胞内部观察的重复单位结构。蛋白质的主体部分经淡化处理，颜色加深的是 N 端延伸部分。VP1 和 VP3 的 N 端分别与对方的主体部分相互作用。而 VP2 在酶切下 VP4 后，剩下的 N 端延伸部分结合在自身亚基的主体部分。右：从内部观察的病毒粒子亚基绕五重轴的折叠。VP1 组成外表面，VP3 的 N 端（红色）嵌入正中轴心的开口。在结合到受体后，VP4（青色）离开病毒粒子并使 VP1 的 N 端暴露出来。

感染开始于病毒粒子对宿主细胞表面受体分子的结合。脊髓灰质炎病毒的受体是一种细胞粘连蛋白 CD155 或 PVR（poliovirus receptor, 脊髓灰质炎病毒受体）。病毒与受体蛋白的相互作用会介导病毒本身发生大的构象变化。已有实验证实在结合到受体后，VP4 会离开病毒粒子，而 VP1 的 N 端区域向外暴露出来。这证明病毒粒子本身是动态的，允许在衣壳上形成开孔以使切割下来的多肽释放出来。VP4 的 N 端对膜的亲和性较强。这些变化很可能就参与了在宿主细胞膜上打开一个通道并使 RNA 进入宿主细胞质的过程。但是人们对这个过程具体的细节尚不清楚。

15.3 病毒对宿主核酸的结合

在组装过程中，将感染性的粒子融合到正确的核酸上是非常重要的。这个过程存在多种机制，但最著名的是噬菌体 MS2 使用的方式。噬菌体 MS2 是一种简单的 T=3 的病毒，它由 180 个拷贝的同种衣壳蛋白和一条 3500bp 的单链 RNA 分子组成。衣壳蛋白能够结合到一段 19bp 的 RNA 发夹结构上，阻遏了复制酶基因起始密码子的识别，从而抑制一个病毒基因的转录。而这个衣壳蛋白与 RNA 的结合同时也是病毒粒子组装的起始步骤。因此 RNA 发夹结构既是一个转录调控子又是病毒核衣壳化的一个信号。

MS2 衣壳蛋白具有由一个折叠片和两个螺旋结构组成的独特的折叠方式。两个单体通过很强的相互作用形成二聚体，该作用来自于一个亚基的螺旋结构嵌入另一个亚基的口袋中（图 15.11）。

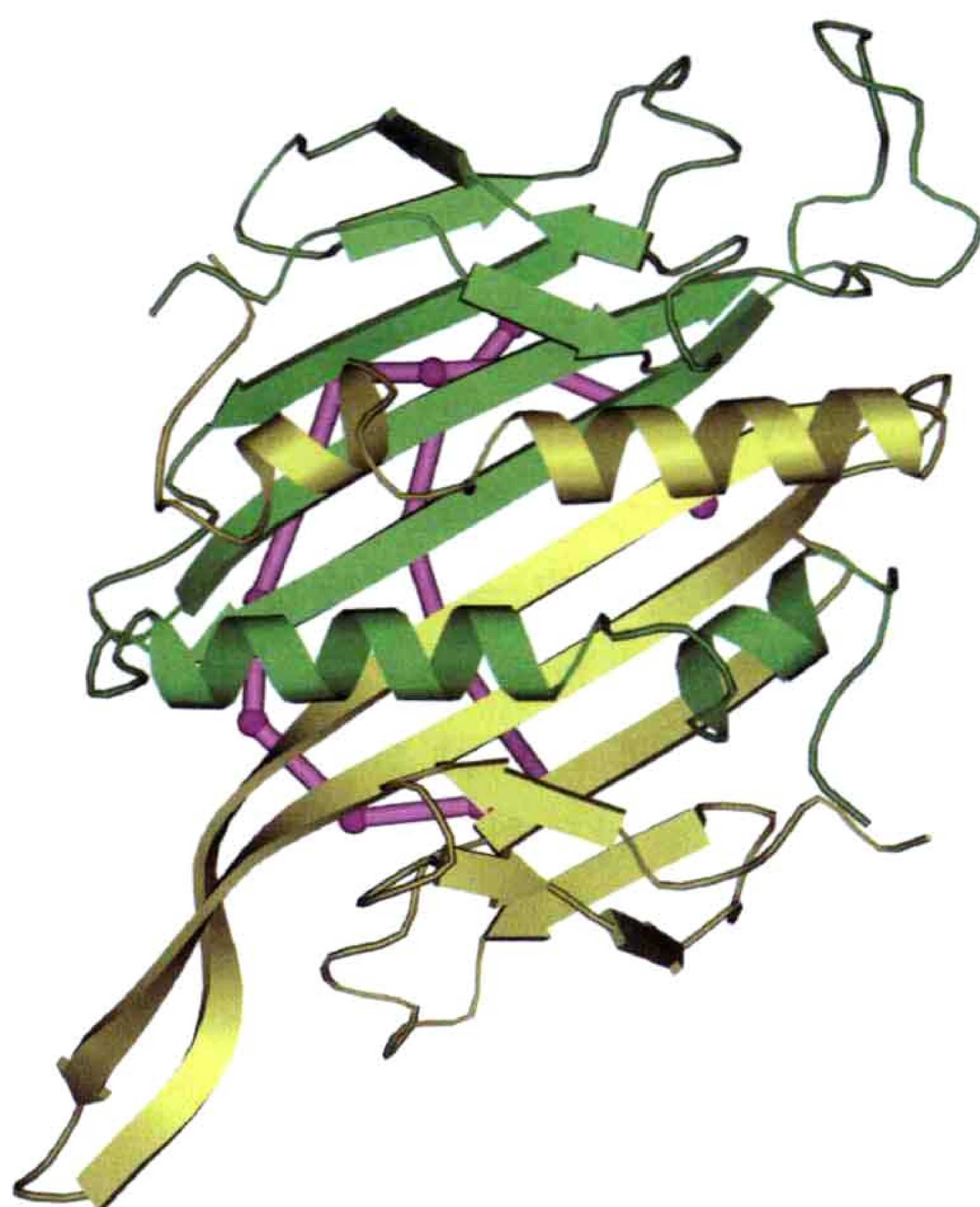


图 15.11　从病毒粒子外部自下而上的方向观察的噬菌体 MS2 衣壳蛋白的结构。紫色表示的是部分结合在蛋白质上的 RNA（PDB：2BU1）。

很多 RNA 结合蛋白都是由反平行折叠片和几个螺旋结构组成的。虽然与此类似，但 MS2 衣壳蛋白与其他的 RNA 结合蛋白的拓扑结构并不相同。MS2 在折叠片的表面处结合 RNA（图 15.12）。MS2 与 RNA 的结合是通过 RNA 分子的单链区实现的，如靠近发夹结构处的四核苷酸环状结构和发夹结构茎上突起的腺苷酸 A。

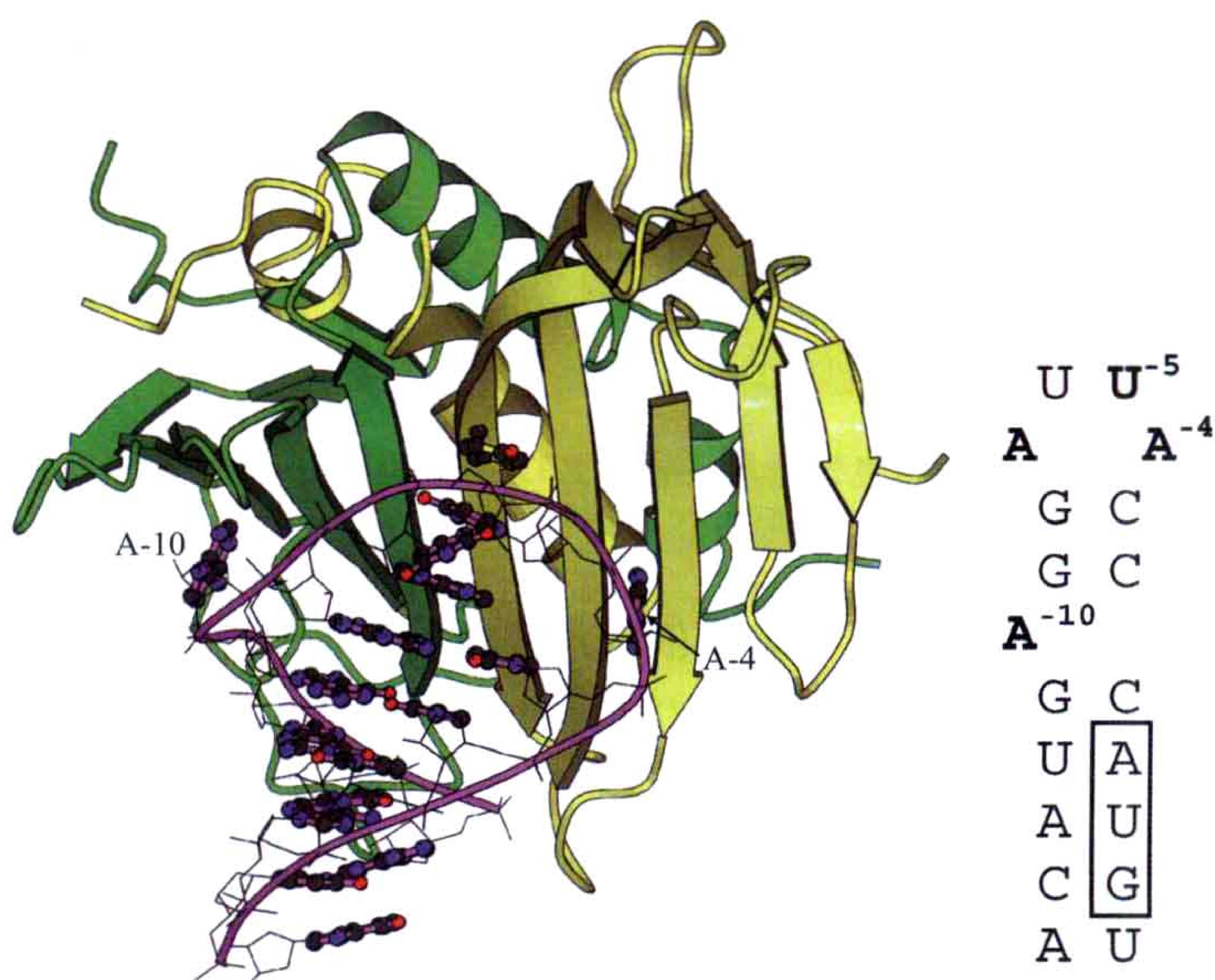


图 15.12　噬菌体 MS2 衣壳蛋白二聚体的结构（PDB：1ZDI）。－ 10 和－ 4 处的腺嘌呤结合在二聚体两个单体各自的口袋中。尿嘧啶堆积在其中一个亚基侧链上的酪氨酸附近。右边的核酸序列表示 RNA 发夹结构的二级结构。框中的 AUG 是复制酶的起始密码子。

15.4 一种复杂病毒——T4 噬菌体的结构

细菌病毒本身并不进入宿主细胞，而是通过不同的机制将自身的基因组注射到细菌细胞内部。上面描述过的小 RNA 病毒利用细菌间常用来传递遗传物质的菌毛（pili）来使自己的 RNA 进入细菌细胞。而另一大类细菌噬菌体自身带有能够注射遗传物质的尾（tail）。带有尾部的噬菌体中最为复杂的一个家族称为肌尾病毒，T4 噬菌体就属于这个家族（图 15.13）。T4 噬菌体的头部为 T=13 的二十面体，另有一圈额外的亚基使之拉长。尾部末端的基板（baseplate）连有两种类型的尾丝（fiber）。基板由至少 14 种不同的蛋白质组成，并具有部分六重对称性。对于噬菌体尾部的结构和功能的理解来自于电镜、单独蛋白或蛋白质复合体的晶体结构及广泛的生化研究的结果（图 15.14 和图 15.15）。

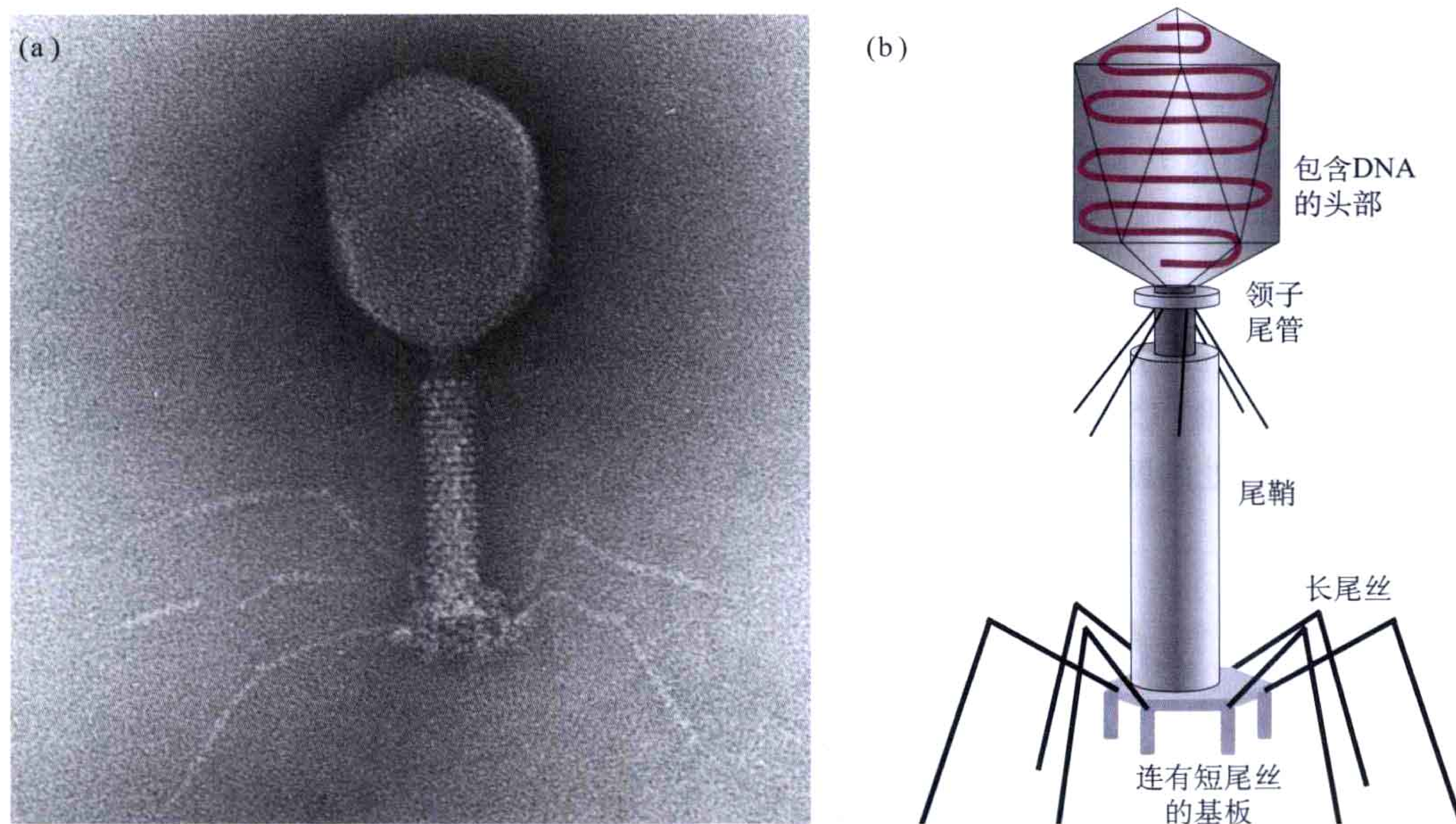


图 15.13 T4 噬菌体。（a）噬菌体电镜照片（感谢 R.Duda,Pittsburgh 友情提供）；（b）病毒粒子主要部分的示意图。

完整病毒粒子的组装按照每个组分依次加入的特别方式进行。头部的组装利用了一个脚架蛋白，该蛋白质在孔道蛋白（portal protein）将 DNA 注入细菌细胞前就会降解。尾部独立组装并连接到包裹了 DNA 的头部。

在成熟的病毒粒子中，尾部就像一个弹簧，短的尾丝与细菌的相互作用会引起基板构象的变化。这些变化诱导尾鞘（tail sheath）收缩。在这个过程中，尾管穿过细胞膜并将 DNA 注射到细胞内部。基板含有具有这个特殊功能的各种蛋白质，其中一个含有可降解细菌细胞壁活性的结构域。这个结构域（图 15.14）与大多在生命体中发现的溶菌酶结构类似。噬菌体基因组本身也编码了一个溶菌酶（T4 溶菌酶），但是这个酶

是在噬菌体粒子要离开细菌时裂解细菌细胞用的。

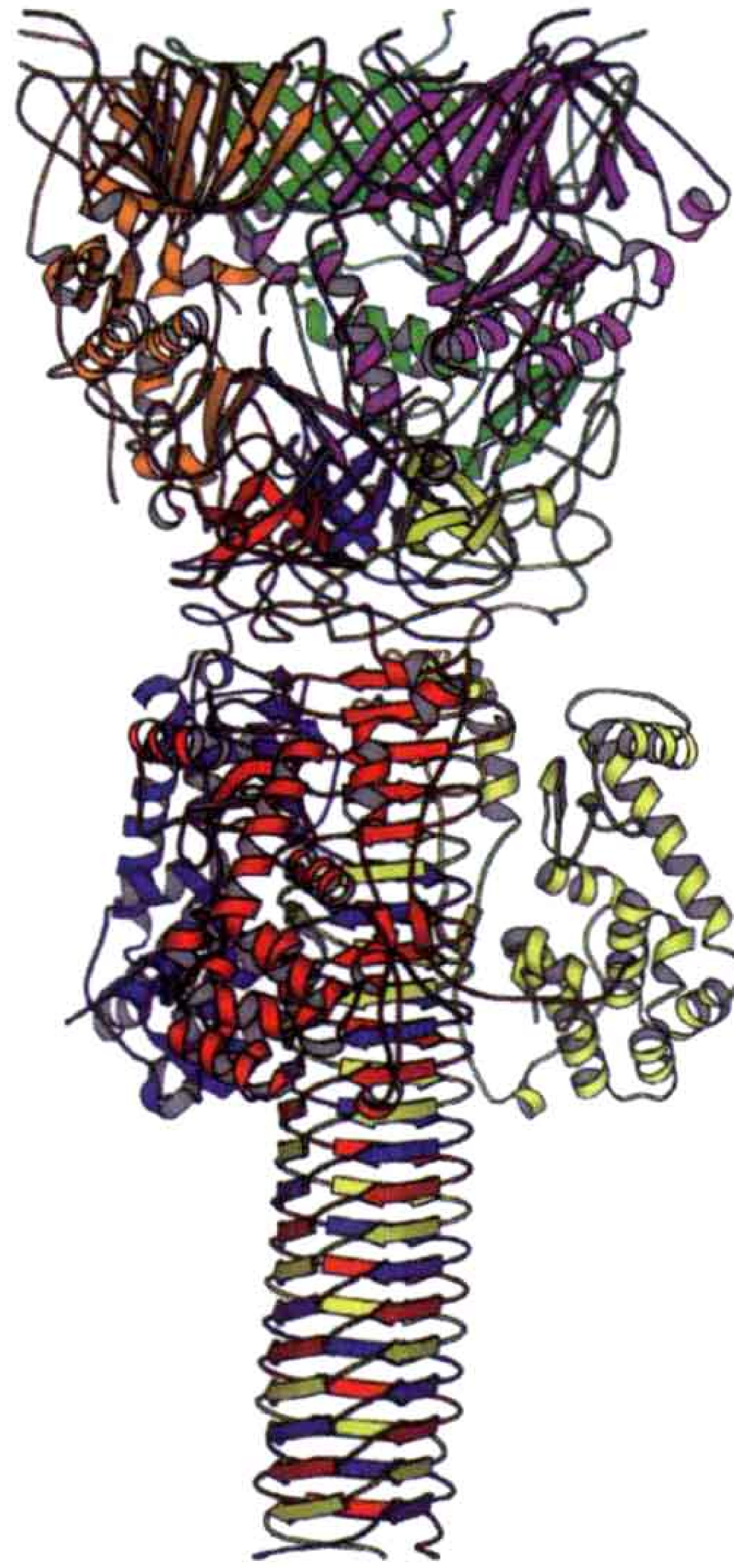


图 15.14　三聚化 gp5-gp21 复合体。gp5 的三个单体分别用红色、蓝色、黄色表示，gp27 的三个单体分别用绿色、棕色、紫色表示。gp5 的溶菌酶结构域位于三聚 β 螺旋的上部并形成整个分子的茎部。

几种参与组成基板的蛋白质的结构已经得到了解析，图 15.15 展示了将详细的晶体结构构建到冷冻电镜图谱的结果。

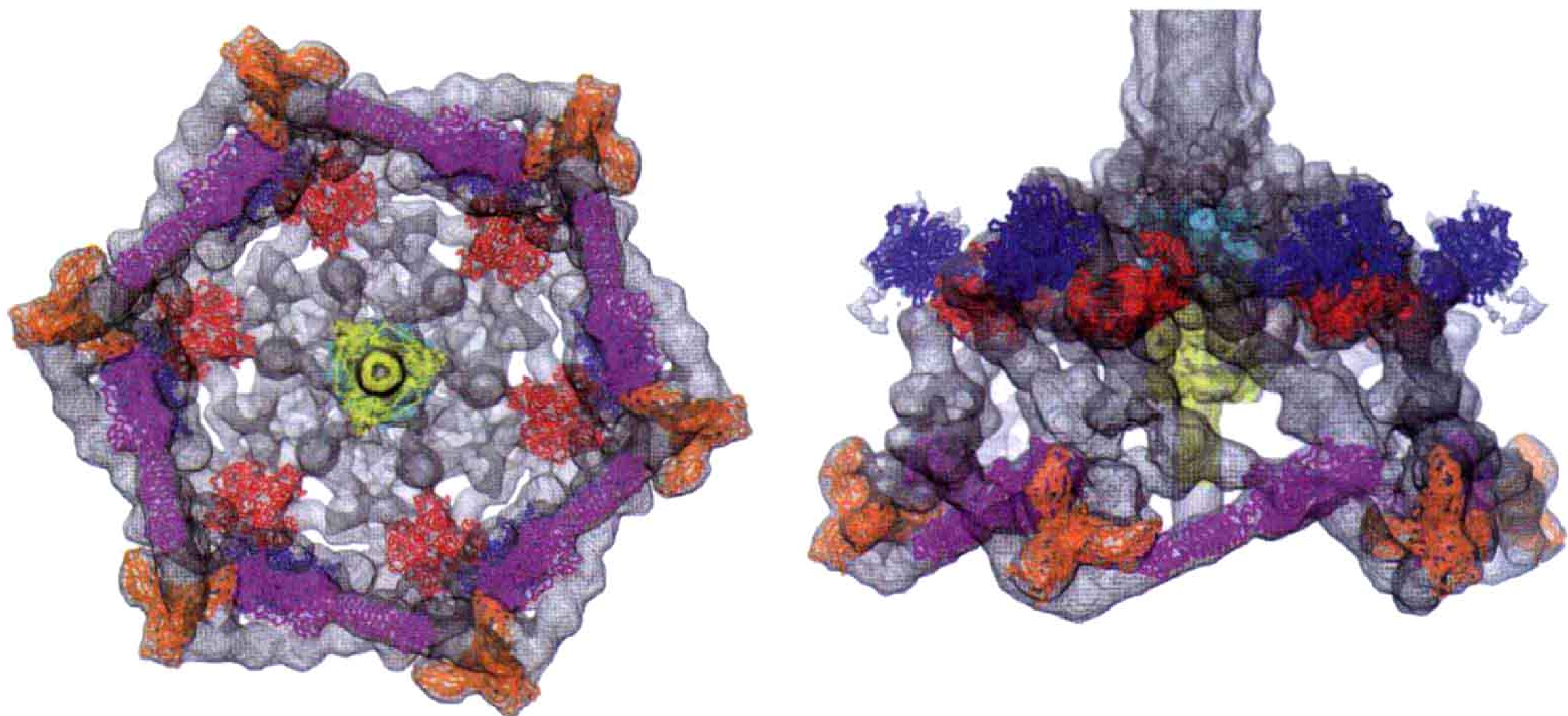


图 15.15　将来自 T4 噬菌体基板的几种蛋白质的晶体结构构建到冷冻电镜图谱的效果图。gp5-gp27 复合体位于模型的中心。gp5 用黄色表示，gp27（几乎不可见）用青色表示。另一个构建到冷冻电镜图谱的蛋白质是 gp9（蓝色）、gp8（红色）、gp11（橙色）和 gp12（紫色）。（感谢来自旧金山大学的 Thomas Goddard 的友情提供）。

延伸阅读

原始文献

Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AGW, Rayment I, Rossmann MG, Suck D, Tsukihara T. (1980) Structure of southern bean mosaic virus at 2.8Å resolution. Nature 286: 33-39.

Bullough PA, Hughson FM, Skehel JJ, Wiley DC. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. Nature 371: 37-43.

Caspar DLD, Klug A. (1962) Physical principles in the construction of regular viruses. Cold Spring Harbor Symp Quant Biol 27: 1-24.

Hogle JM, Chow M, Filman DJ. (1985) Three-dimensional structure of poliovirus at 2.9Å resolution. Science 229: 1358-1365.

Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. Nature 371: 623-626.

Wilson IA, Skehel JJ, Wiley DC. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3Å resolution. Nature 289: 366-373.

综述文章

Hogle JM. (2002) Poliovirus cell entry: Common structural themes in viral cell entry pathways. Annu Rev Microbiol 56: 677-702.

Johnson JE, Speir JA. (1997) Quasi-equivalent viruses: A paradigm for protein assemblies. J Mol Biol 269: 665-675.

Smith AE, Helenius A. (2004) How viruses enter animal cells. Science 304: 237-242.

(邢　栋 译　苏晓东 校)

第十六章
结构生物学和生物大分子的演化

16.1 介绍

18 世纪中叶，法国博物学家布丰（George-Louis Leclerc de Buffon）观察到，所有哺乳动物的肢骨虽然发挥的作用各不相同，但十分相似。灵长类用它们的前肢使用物品，狗和猫用它们的前肢奔跑，而鲸鱼用它们的前肢游泳。布丰也注意到，猪的脚趾从来不接触地面，因此看似没有什么功能。这些观察到的现象让布丰很难用那时的常识来解释，即所有的生物都是被神创造出来的。如果是这样，那么为什么猪的肢体显然是不完美的呢？在布丰的《动物的自然史》一文中，他提出哺乳动物的肢骨可能来自于同一祖先。看似无用的猪的脚趾可能遗传自完好的能发挥作用的祖先的脚趾。布丰的学生拉马克（Jean-Baptiste de Lamarck）进一步发展了“共同祖先起源”的思想，并尝试着提出了一个普遍的机制来解释为什么改变能够发生。他最重要的贡献是巩固了“生物时刻都在发生变化”这一思想。

19 世纪时，地理学家积累的证据表明，地球已经存在并演化超过百万年，而并非数千年（图 16.1）。这是达尔文（Charles Darwin）和华莱士（Alfred Russel Wallace）独立提出新的进化理论的主要背景之一。达尔文的理论基于一些观察到的现象，其中最重要的现象是以下三点。

• 每种生物都能迅速繁衍并覆盖地球。

• 在一种生物内，每个个体之间都是不同的。

• 后代与父母是相似的，因为它们遗传了父母的特征。

1858 年，达尔文在他的“物种起源”里阐释了个体之间的不同在个体生存和繁殖中产生怎样的影响。因此遗传下来的特性能从一代传到下一代。假设任何一种特性能够提高其携带者生存和繁殖的概率，那么这种特性将成功地传递给后代。



图 16.1　来自瑞典 Dalarna 省 Dalhalla 市的三叶虫化石（Lennart Edvardsson 友情提供）。化石在推断高等生物进化史的过程中发挥了关键作用。而遗憾的是，并不存在生物大分子的化石记录。

现在你可以尝试将上文中提到的“生物”替换为“生物大分子”，如酶，这样你将了解生物大分子进化史背后的基本观点。相同生物大分子的物理、化学和生物学性质在相同物种的不同个体中会有不同。生物大分子性质的各种改变如果促进生物体的存活和繁殖，就会成功地传递给后代。换句话说，所有现存的蛋白质对编码它的生物体已经有了有益的影响或尚存有益的影响。然而，这些蛋白质中的一些就像猪的脚趾。我们将会在后面的章节中详述这些，那时我们将集中在蛋白质进化的具体例子上。

19 世纪中叶，摩拉维亚修道士孟德尔（Gregor Johann Mendel）对一系列植物进行了杂交实验。他的工作证实了遗传的颗粒性。孟德尔也建立了用于预测遗传特性从双亲到子代传递过程中的统计法则。然而，达尔文和孟德尔并不知道，数十年后，他们的发现注定非常紧密地联系在一起。

16.2 突变

16.2.1 分子钟

核酸分子是进化和遗传过程的关键分子。它们储存有遗传信息并把遗传信息转化为表型特征，也把遗传信息从一代传到另一代。它们同样记录了进化的历史。我们只需要了解怎样从这些信息中解读积累的历史。

DNA 序列在复制过程中几乎是完全精确的拷贝。然而，DNA 复制和修复过程中，错误还是会以很低的频率发生，进而产生出新的序列。这些错误被称为突变，它们是达尔文所观察到的变异现象的分子机理。突变也是进化过程中新物种产生的源泉。在以下的讨论中，我们将集中阐述每种突变类型对遗传物质及其产物，以及对进化产生的影响。

突变通常被认为是随机发生的，并与发生突变的生物体的适应性相关。自发产生突变的频率很难测量，只能间接估算。在哺乳动物细胞核中，自发突变的频率是每年每个核苷酸产生 $3\times10^{-9}\sim5\times10^{-9}$ 次替换，但是在微生物中，这个频率大大提高。简而言之，物种与物种之间及基因组的不同位置之间，突变频率差异极大。基因组的一些位置相比于其他位置来说更倾向于发生突变，这些位置被称为热点。

许多突变不改变基因编码的蛋白质序列。这是因为通用密码子中，大部分氨基酸可由多个密码子（碱基三联体）编码。不改变氨基酸的突变称为沉默突变或同义突变。这些突变不影响相应的蛋白质功能，因此不影响其所在的生物体的存活与繁殖。因此，沉默突变不太可能受到自然选择的影响。改变氨基酸序列的基因突变称为非同义突变。许多氨基酸的改变并不改变蛋白质的结构和功能特性。因此，一些非同义突变对于其所在的生物体来说可能非常近似于中性。

发生于个体水平的自然选择被定义为：遗传背景不同的个体所具有的不同的繁殖能力。换句话说，拥有不同遗传背景（基因型）的个体具有不同的特征，这些特征在特定环境条件下可能是有益的、有害的或者中性的。如果这一系列特征是有害的，那么繁殖的成功率将会减小，基因型也不会成功传递给后代。一个基因型的适应性是个体生存和繁殖能力的量度。在某些极端稀少的情况下，突变会增加携带该突变的个体的适应性。这些突变被认为是有益的并受到正向选择。注意以下几点非常重要。

（1）任何生物大分子的突变形式在某些环境条件下都可以是有害的，并因此对个体产生负面影响；而在另一些不同条件下，它可以是有益的。

（2）实验室中被认为是有益的分子特性不一定对生物体来说是有益的。

1968 年木村资生（Motoo Kimura）提出了分子进化的中性学说。在分子水平上，大部分突变是中性的，它们既不赋予携带者有益的特性，也不赋予有害的特性。这些突变中的一部分在种群中传播开来并被固定下来。我们可以假设中性突变被固定的频率是恒定的，是总的突变频率的一个反映，如中性突变频率可以作为一个分子钟。因此，如果大部分突变不影响蛋白质的功能，那么在各种生物体中，子代蛋白都应该以非常相似的速率积累突变。因此，不同生物体的某一特定蛋白的进化速率随时间的改变应该是恒定的。我们应该怎样推断突变积累的速率，或者说我们怎样测定分子钟的速率呢?

16.2.2　序列分析告诉了我们进化过程

分子进化的研究集中在如下两个子课题上（它们实际上是有重叠的）。

• 重建基因、基因组和生物进化的历史。
• 理解生物大分子的进化。

每个基因和相应的蛋白质在从其祖先进化的过程中已经积累了很多突变。核苷酸和氨基酸序列因此从某种意义上讲，是历史记录的积累，并能用于推断蛋白质及编码该蛋白质的生物体的进化历史。我们想要了解基因、蛋白质，或者生物随时间的改变、改变速率及改变机制。换句话说，对于某一蛋白质，我们想要了解它 100 万、1000 万、1 亿或者 10 亿年前的样子，以及它的性质是以什么样的速率改变的，我们也想了解什么分子机制可能促进了蛋白质的改变。

为了研究某一蛋白质的分子进化，我们必须首先测定该蛋白质精确的一级结构或者其相应基因的一级结构。我们也需要测定其他物种中同源蛋白（homologous）的序列，即这些蛋白质从一个祖先进化而来。最近，从细菌到人类的一系列物种全基因组序列的测定，提供了一个强有力的工具，使我们可以汇集所想研究的几乎任何蛋白质的信息。一旦序列被测定了，它们就可以用于比对。我们想要测定任何改变所发生的位点和数量，如替换、删除、插入，这些改变在祖先分化出的各个物种中已经发生了。然后这些数据可以通过生物信息工具转化为进化速率，可能的中间形态也可得以重建。

形成新物种时，那些从同一祖先进化而来的蛋白质被称为直系同源（orthologs）。即使这些蛋白质是同源的，并拥有相同的祖先，但它们在其各自的生物体中的功能可能不同。另外在蛋白质进化的历史中，编码它们的生物体可以暴露于不同的环境、选择压下或拥有不同的需求。因此，同源蛋白之间的进化速率可能差异巨大。

然而，序列分析的数据必须伴有全面的分子研究来确定一个推测出来的同源蛋白的功能。例如，一个蛋白质的结构或者折叠方式在进化过程中比它的一级序列保守得多。这可以用来帮助认明一个蛋白质的同源蛋白，即使其序列的一致性超出了探查下限。

下一节中，我们将举两个例子来阐述分子进化：新蛋白通过基因复制方式的起源，以及球蛋白物理化学特性的进化。

16.3 新酶的起源

16.3.1 基因加倍

新基因和蛋白质是什么时候、怎样出现的呢？在每个生物体中，新基因每时每刻都在产生或已经产生出来，其中至少有以下 4 种可能的机制来解释新基因的起源。

- 现存基因的加倍。
- 不同基因的一部分组合起来形成嵌合基因。
- 从非编码 DNA 中产生全新的基因。
- 物种之间基因的平行转移。

所有这 4 种机制的例证都有详细的文献记录。迄今为止创造新基因最普遍的方式是通过基因加倍。在每个基因组中，基因都在持续不断的加倍，或者单个片段加倍，或者全基因组的一部分加倍。一对加倍了的基因，即旁系同源基因（paralogs），会面临几个不同的结果。一个拷贝可能因为积累了有害的突变而从基因组中丢失；或者如果这两个拷贝所经历的达尔文式选择是正向的话，它们都能够被保留下来。其中一个拷贝可能获得新的功能（neofunctionalization），或者旧的功能可能被新的拷贝所细分（subfunctionalization），这些新的拷贝的功能将变得更加专一化。在基因获得新功能的情况下，当两个拷贝中的一个拷贝获得了有益的突变而另一个拷贝保留有初始的基本功能时，两个拷贝都倾向于保留下来。与之相反，当两个拷贝由于突变而使功能退化到某一程度，即它们必须共同表达来完成祖先基因的基本功能时，它们也都会因为功能细分而保留。

在接下来的讨论中，我们将集中在参与核酸前体代谢的基因和酶怎样通过基因加倍进化产生：①新的功能，如催化一个不同的反应；②分割初始的功能，如通过“专

一化”或使加倍了的基因对底物的识别范围受到约束。

核酸，即 DNA 和 RNA，是所有细胞的核心大分子。细胞采用多种策略——从头合成（图 16.2）、补救途径和代谢——来保证核酸前体的供给平衡，如嘧啶、嘌呤、核苷及核苷酸。因为这些途径决定了 DNA 和 RNA 的前体池，所以它们直接与遗传物质的复制和基因表达相关。数十年前，研究者们提出，池的不平衡会引发不同的健康问题。迄今为止，从细菌到人类，多个不同生物体的这些途径已得到研究。现在，比较不同生物体中功能相似的酶，可以帮助我们重建进化的中间体，了解这些酶重塑的机制。

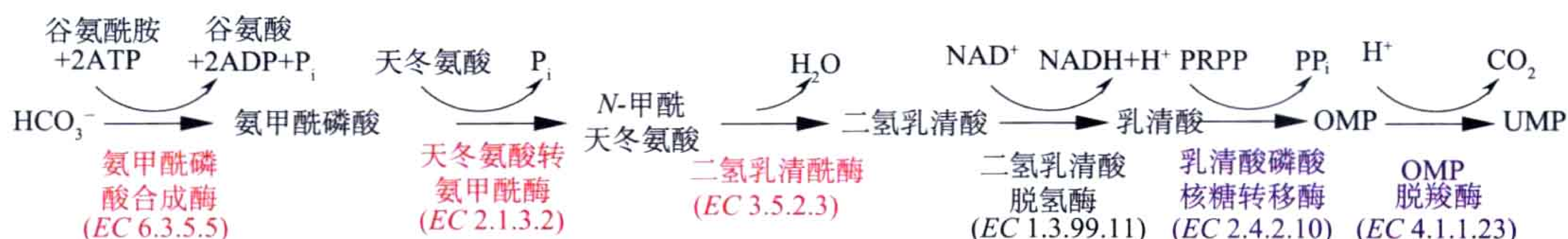


图 16.2　嘧啶的从头合成过程。具有多种酶活的蛋白质 CAD 的催化活性以红色表示（它在高等生物中催化前三个反应，如动物中）；UMP 合酶的催化活性以蓝色表示（PRPP: 5- 磷酸核糖 -1- 焦磷酸；OMP: 单磷酸乳清酸；UMP: 尿嘧啶单磷酸）。

16.3.2　脱氧核糖核苷酸的补救途径：功能细分

细胞需要持续供给脱氧核苷三磷酸来用作遗传物质复制和修复时的原材料。DNA 前体可通过从头合成和补救途径来供给。在从头合成途径中（图 16.2），这些前体物质通过核苷酸还原酶 RNR（见 5.2 节）还原核苷酸来产生。脱氧核苷激酶催化脱氧核苷（dN）磷酸化为相应的脱氧核苷单磷酸（dNMP），这个酶是补救利用胞外产生的脱氧核苷或胞内 DNA 降解产生的脱氧核苷时的关键酶。随后，dNMP 分别经脱氧核苷单磷酸激酶和脱氧核苷二磷酸激酶催化，磷酸化为二磷酸（dNDP）和三磷酸（dNTP）。在人体内，补救途径也对激活一些药物前体有极其重要的作用，如抗病毒药物 acyclovir 和 AZT（见第三章），以及抗癌药物 gemcitabine。

脱氧核苷激酶（图 16.3）在所研究的不同物种中显示出很大的多样性。哺乳动物中有 4 种脱氧核苷激酶，它们的特异性有部分重叠：胞质的胸腺嘧啶脱氧核苷激酶 1（TK1）只磷酸化胸腺嘧啶脱氧核苷；线粒体的胸腺嘧啶脱氧核苷激酶 2（TK2）磷酸化胸腺嘧啶脱氧核苷和胞嘧啶脱氧核苷；胞质的胞嘧啶脱氧核苷激酶（dCK）磷酸化腺嘌呤脱氧核苷、鸟嘌呤脱氧核苷和胞嘧啶脱氧核苷；线粒体的鸟嘌呤脱氧核苷激酶（dGK）只磷酸化嘌呤类脱氧核苷，即腺嘌呤脱氧核苷和鸟嘌呤脱氧核苷。在昆虫体内，只有一种脱氧核苷激酶，它能催化所有 4 种天然底物，并与人类的 TK2 相关（图 16.3）。

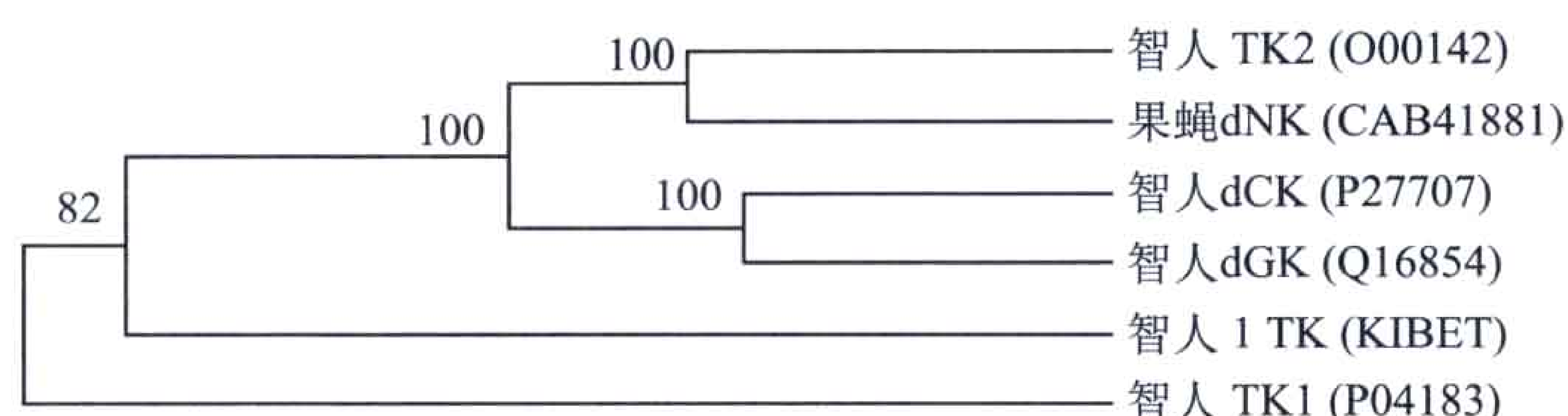


图 16.3　4 种人类的脱氧核苷激酶 dCK、dGK、TK1 和 TK2，一种可催化多底物的昆虫 DmdNK，以及单纯疱疹病毒 1（HSV-1）的 TK（GenBank 的登录号在圆括号内）。三种人类的激酶 dCK、dGK 和 TK2 属于同一个家族，通过至少两轮的基因加倍过程起源于同一祖先；而人类 TK1 很明显有一个不同的起源，而且从进化角度看与其他三种酶无关。

包括单纯疱疹病毒 1 胸腺嘧啶脱氧核苷激酶 HSV1 TK（它激活抗病毒药物 acyclovir）在内的几个病毒的激酶也已经得到研究。HSV1 TK 是第一个被解析了三维结构的脱氧核苷激酶。然而，这个酶比人类的激酶大，且与人类的激酶亲缘关系并不紧密。

果蝇（*Drosophila melanogaster*）的脱氧核苷激酶（图 16.4）DmdNK，拥有很高的催化速率和广谱的底物特异性，它能磷酸化所有 4 种天然的脱氧核苷和一些医疗上重要的类似物。DmdNK 是第一个被解析结构的真核生物激酶，它提供了一个解释动物中激酶结构 - 功能普遍特性的初始模型。虽然 DmdNK 和人类 dGK 与 HSV1 TK 的氨基酸序列只有少于 10% 的一致性，但这三种酶都有相似的折叠方式。DmdNK 和 dGK 的折叠方式由 8 ~ 10 个 α 螺旋围绕 5 束平行的 β 折叠片构成（图 16.4）。这些蛋白质都属于拥有一个可结合核苷酸的 P- 环（P-loop）的蛋白质大类群（激酶，G 蛋白及其他）。令人惊奇的是，与 HSV1 TK 相比，这个折叠方式更类似于一种单磷酸激酶，即胸腺嘧啶脱氧核苷单磷酸激酶，它催化补救途径中的第二步反应。这个发现清楚地表明，脱氧核苷激酶和脱氧核苷单磷酸激酶具有相同的起源，它们可能来自于一个两种反应都能催化的原酶。基于加倍过程，一个酶系成为专门催化脱氧核苷底物的酶系，而另一个酶系成为专门催化脱氧核苷单磷酸底物的酶系。

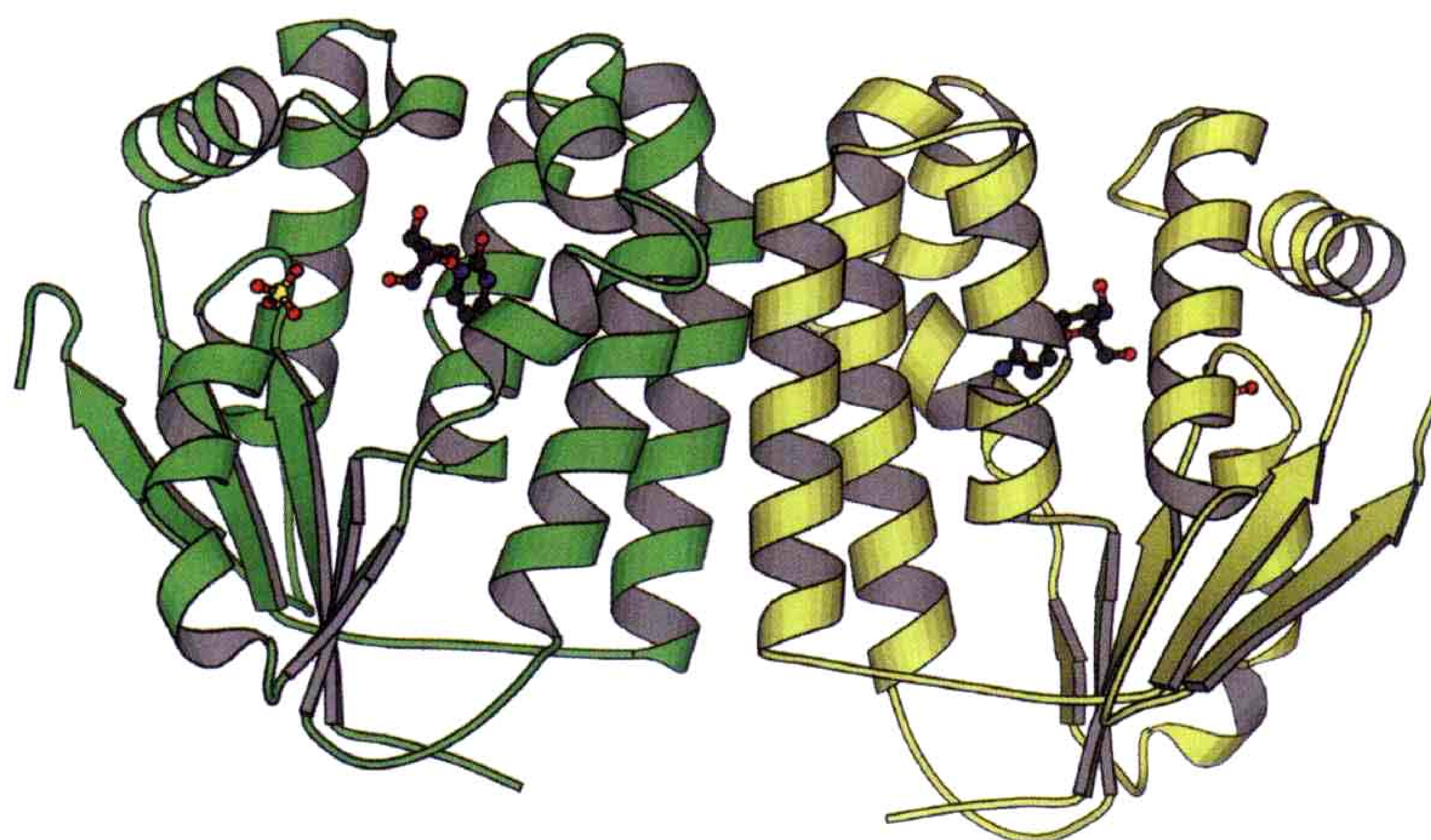


图 16.4　果蝇脱氧核苷激酶的一个二聚体。它是一种可磷酸化所有 4 种底物的酶，看起来像现代哺乳动物激酶的一个祖先，或者是退化的一个例证？（PDB: 1J90）。

与之相反，人类TK1与其他激酶的亲缘关系较远（图16.3）。它具有一个不同的折叠，N端为α/β结构域，C端是一个较小的结合有一个锌离子的结构域。它只有在结合磷酸供体的P-环处与其他脱氧核苷激酶具有序列相似性（图16.5）。

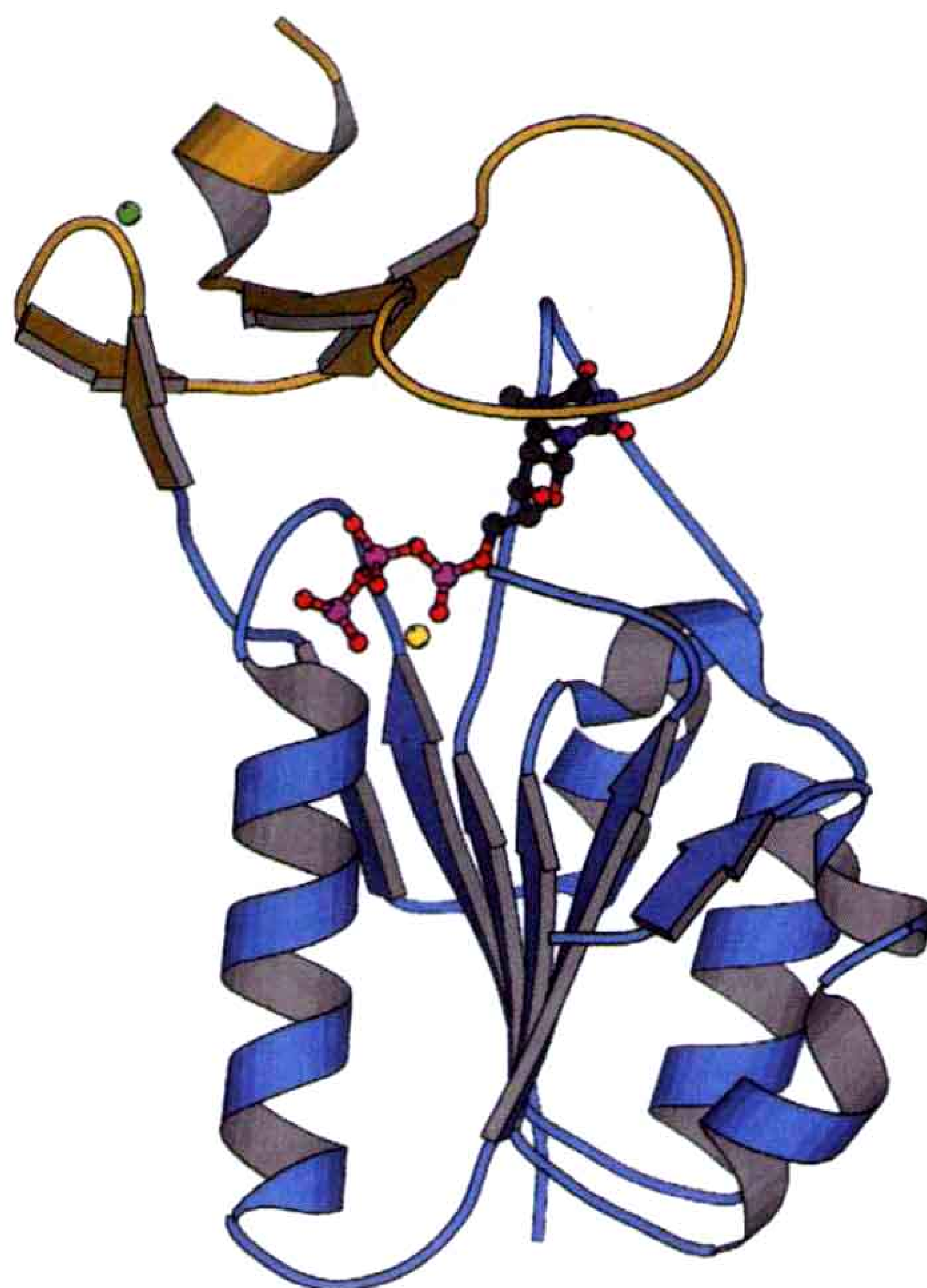


图16.5　人类胸腺嘧啶脱氧核苷激酶1。TK1有两个结构域：一个中心有6束平行折叠片的α/β结构域和一个较小的结合锌离子的结构域。这个蛋白质与dCK、dGK和DmdNK激酶相比，具有一个不同的折叠方式（PDB: 1XBT）。

α/β结构域有一个位于中心的6束平行的β折叠片，这个β折叠片由位于一侧的较长的α螺旋和一个卷曲，以及位于另一侧的三个较短的螺旋所围绕。整个α/β结构域在折叠方式上实际更类似于RecA和F1 ATP酶的ATP结合结构域。类似的结构域同样也在解旋酶和DNA修复蛋白中被发现，这些蛋白质能够与核苷三磷酸的水解相偶联，解开DNA或RNA双链。此处我们看到的是趋异进化的一个例证——一个原始基因的加倍导致现在催化完全不同反应的酶的产生。

令人困惑的是为什么自然界会进化出完全不同的两种激酶家族，即TK1和dCK/dGK/TK2，来催化一系列非常相似的反应。尤为困惑的是结构上并不相关的TK1和TK2两个酶的底物特异性却非常相似，这两个酶都以胸腺嘧啶脱氧核苷作为主要的底物，磷酸化胸腺嘧啶脱氧核苷。

为研究属于dCK/dGK/TK2家族的激酶当前特性的起源，它们的结构-功能关系已经得到研究。结构比对发现了DmdNK和dGK中的一些关键氨基酸残基，这些残基参与其与脱氧核苷底物的直接接触。与具有有限底物特异性的人类的酶相比，DmdNK具有一个更宽的底物结合裂隙，因而这个裂隙更有可能赋予了该酶多底物的特性。突变实验清楚地表明，围绕在底物嘧啶环周围的Val84、Met88和Ala110三个残基，对于DmdNK的底物特异性起关键作用（图16.6）。这三个位点在dCK、dGK和TK2中为

不同的氨基酸残基所占据，而这些围绕在底物周围的残基在所有激酶中非常保守。

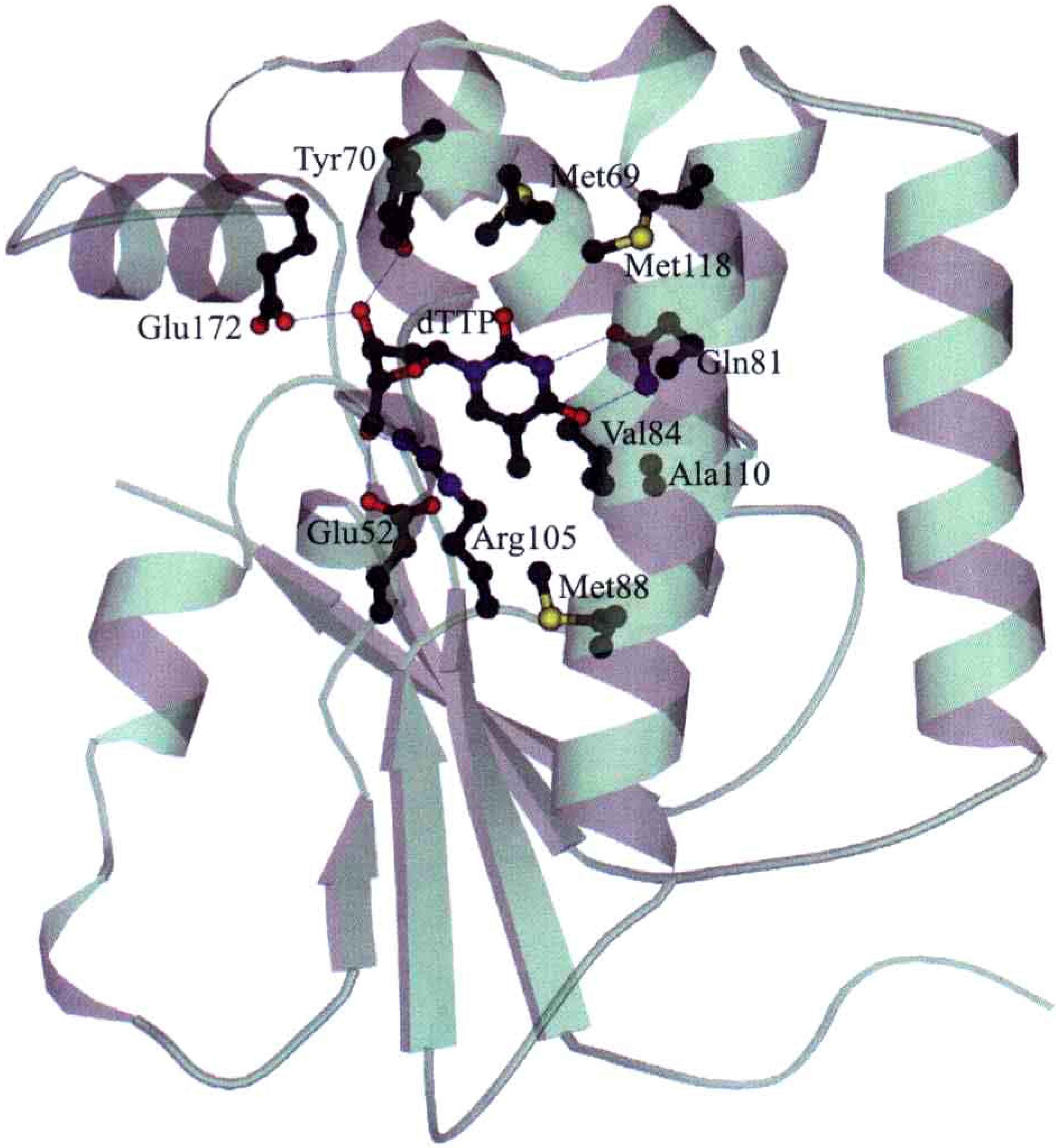


图 16.6 在 dCK、dGK、TK2 和 DmdNK 脱氧核苷激酶中，底物的特异性主要由仅仅三个氨基酸残基决定，在 DmdNK 中它们的位置为 84、88 和 110（PDB: 1J90）。

研究者观察到，当 DmdNK 和 dCK 中的这三个残基被其他激酶中相应的残基所取代时，底物的特异性能发生巨大的变化。仅仅通过改变这三个位点，DmdNK 可以从一个多底物酶变为一个更专一于一个底物的酶。与之相似，当 dCK 中相应的三个残基（Ala100、Arg104 和 Asp133）进行了模拟 DmdNK 形式的突变后，与野生型的 dCK 相比，这个三突变的 dCK 对 dC 的催化效率增高，这一特点和 DmdNK 相似。关于现代激酶的起源问题，这些结果告诉了我们什么？

属于 dCK/DmdNK/dGK/TK2 组（图 16.3）的动物激酶有一个共同的起源，但当前的酶具有相差甚大的底物特异性。很明显，只有几个氨基酸残基赋予了底物的特异性。因此，具有催化多底物特性的原始的酶的进化历史可能如下：基于数轮基因加倍过程，新的激酶形式从动物系中起源。为使这些酶具有各自的特异性，只需要有限数目的突变。原始的具有广谱底物特性的酶分化为三个酶系，导致了当前只能催化窄谱底物的酶的产生。而最初研究者认为 DmdNK 代表了一个活化石，如一个古代的多底物特性的酶，现在研究者确信，这一特性在昆虫中具有更接近现代的起源，一个底物相对特异的类似 TK2 的酶经历了退化过程并使自身底物特异性降低。最近的点突变实验表明，仅仅通过改变少数几个氨基酸残基就可以使最初类似 TK2 的酶进化为当前具有多底物特性的酶。然而，在生物体水平上这个进化途径产生的原因目前还不清楚。

16.3.3　嘧啶的分解代谢：新功能的产生

细胞内合成的大部分中心代谢物也必须通过降解来确保细胞池的平衡。很明显，真核生物中嘌呤和嘧啶的水平如果不受调控，则是具有毒性的。嘧啶的分解代谢途径对于癌症患者极为重要，因为该途径可降解一些化疗药物；该途径在植物、单细胞真核生物和细菌中也很重要，因为它可降解基于嘧啶环的生物杀灭剂。在人体中，该途径有三种酶参与：由 *PYD1*（或 *DPYD*）基因编码的二氢尿嘧啶脱氢酶（DPDase）、由 *PYD2* 基因编码的二氢尿嘧啶酰胺水解酶（DHPase），由 *PYD3* 编码的β- 丙氨酸合酶（缩写为 BSase 或 bASase 或 bUPase）（图 16.7）。

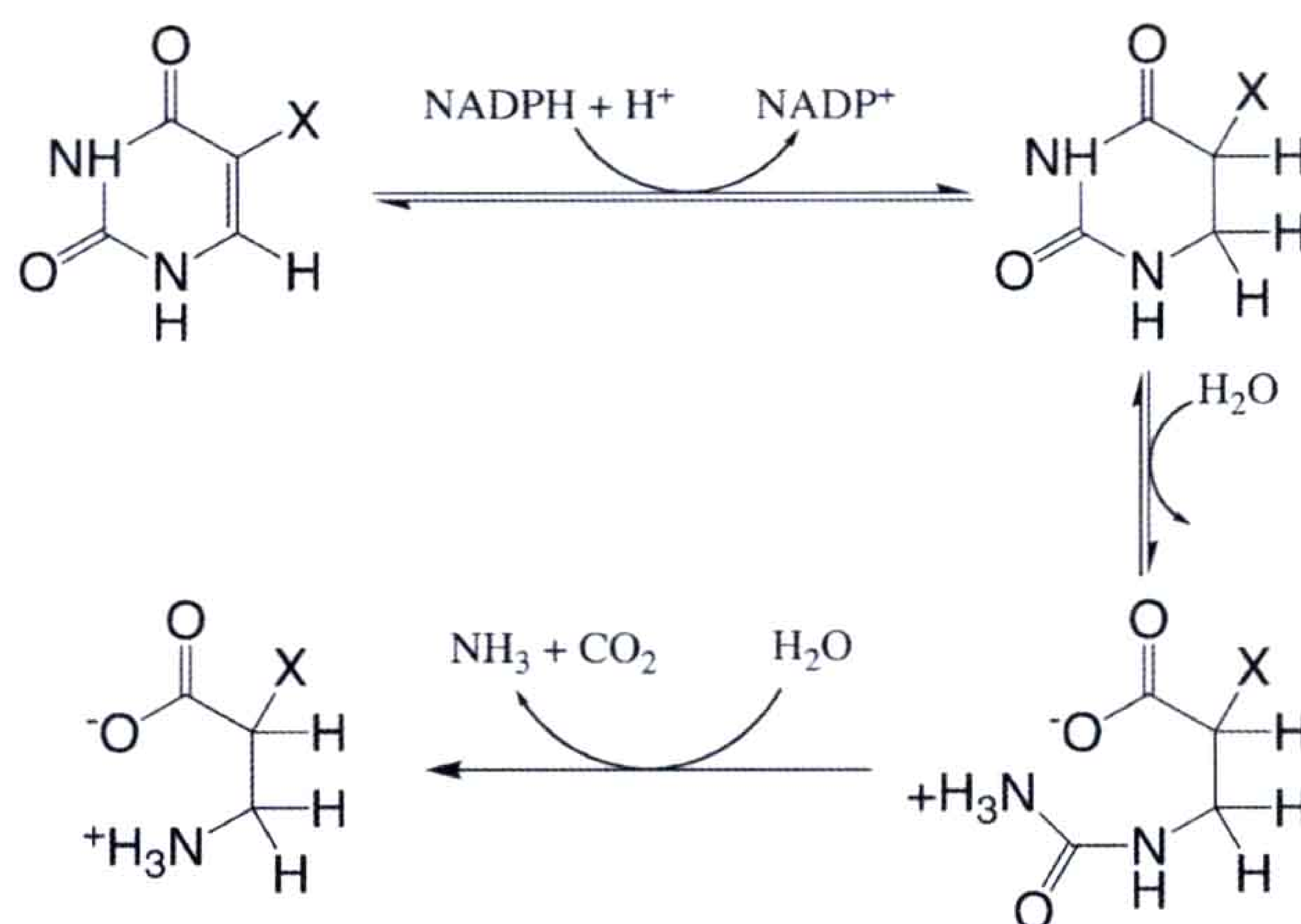


图 16.7　存在于哺乳动物中的嘧啶的还原降解。X 代表 H（在尿嘧啶中）或代表一个甲基基团（在胸腺嘧啶中）。注意，这三步酶促反应和中间体分别类似于生物体中嘧啶从头合成途径的第四、第三和第二步（图 16.2）。简而言之，尿嘧啶先通过二氢尿嘧啶脱氢酶（DPDase, EC 1.3.1.1）降解为二氢尿嘧啶，再通过二氢尿嘧啶酰胺水解酶（DHPase, EC 3.5.2.2）降解为β- 脲基丙酸，然后通过β- 丙氨酸合酶（BSase, EC 3.5.1.6）降解为β- 丙氨酸。

这些参与分解代谢的酶在功能上分别类似于二氢乳清酸脱氢酶（DHODase）、二氢乳清酸酶（DHOase）、天冬氨酸转氨甲酰基酶（ATCase），这些酶分别催化嘧啶从头合成途径中的第四、第三和第二步反应（图 16.2）。这并不令人感到惊奇，因为这些合成代谢和相应的分解代谢的酶在本质上催化的是同样的反应，但反应的方向相反。很显然，在进化过程的某一点上，属于从头合成途径中的基因加倍了，并获得了一个新的功能。或者，这两类反应都能被原始的酶所催化，而反应的方向由底物的浓度决定。人类 DPDase 的基因中包含一个片段，这个片段与编码合成代谢的 DHODase 的基因同源。而 DHOase 和 DHPase 的相似性甚至更加明显（图 16.8）。当原始的可能催化一个可逆反应的 DHOase 的编码基因加倍后，一个拷贝变为专门催化合成代谢的酶，而另一个拷贝变为专门催化分解代谢的酶。这两个酶仍然保留包含两个锌原子的非常相似的反应中心。我们可以推测，只需要改变少数几个氨基酸残基就可以使合成代谢的酶变为分解代谢的酶。

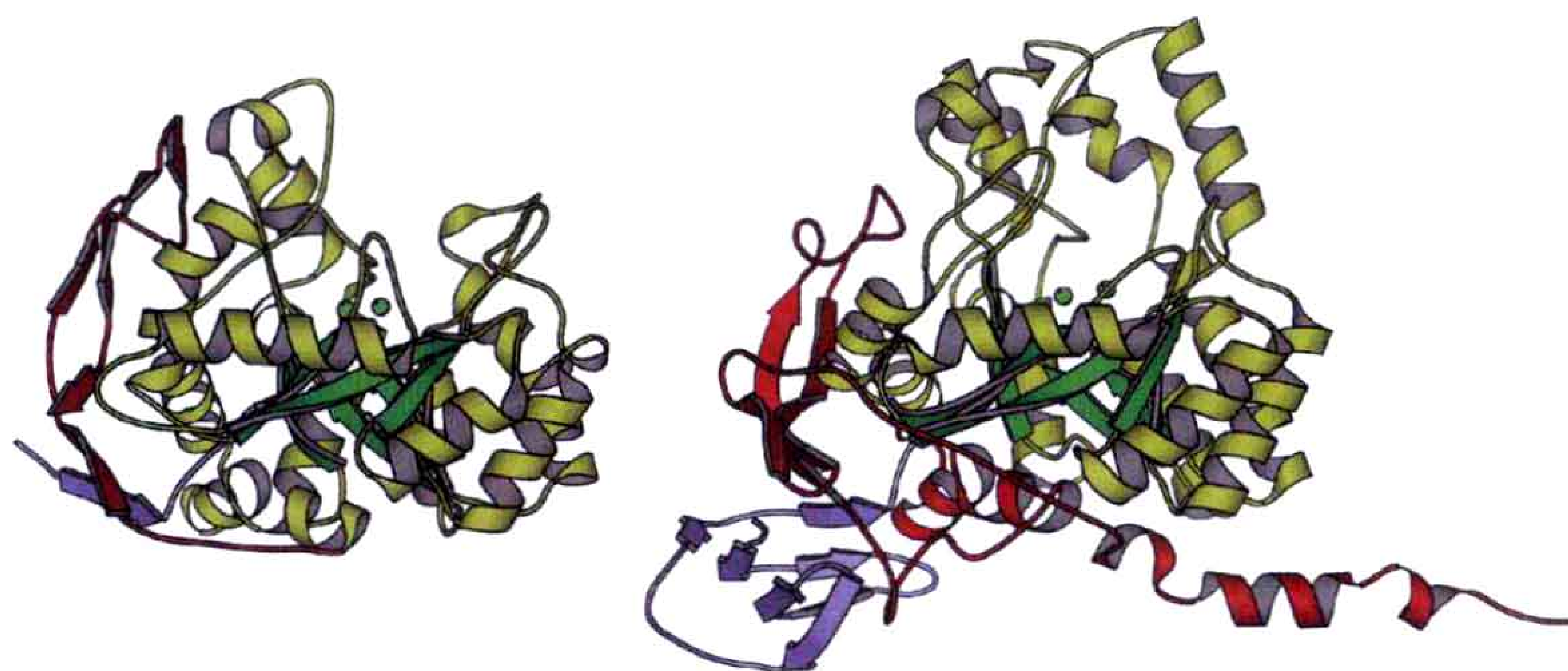


图 16.8　催化合成代谢的 DHOase（左，PDB: 1J79）和催化分解代谢的 DHPase（右，PDB: 2FTW）折叠方式非常相似，暗示它们通过一个原始基因的加倍起源于同一祖先。

酵母中存在两个拷贝的 DHOase：一个是有活性的，由 *URA4* 基因编码；另一个是沉默的，由 *URA2* 基因的一部分编码。令人惊讶的是，这个有活性的酵母基因在进化树上自立门户。这个基因可以代表一个“类似祖先”的基因（或者一个活化石），它可能已经被改变为更专一的“现代”形式。另外，沉默的 DHOase 拷贝看似与真核生物的 DHOase 紧密相关。很明显，这个拷贝在进化过程晚期的某一点被关闭了，而那个“类似祖先”的基因成为高效的现代酶。

第三步催化由 BSase 完成，它由 *PYD3* 基因编码。正常的酶在几乎所有真核生物中都有发现，但在酵母中真正的 *PYD3* 基因很明显丢失了。在酵母中，一个类似外肽酶的酶被招募进来，并在随后，其底物的特异性被改变——β - 脲基丙酸成为被选择的底物。虽然 BSase 和与外肽酶类似的酶之间的序列一致性非常低，但它们的折叠方式非常相似（图 16.9）。

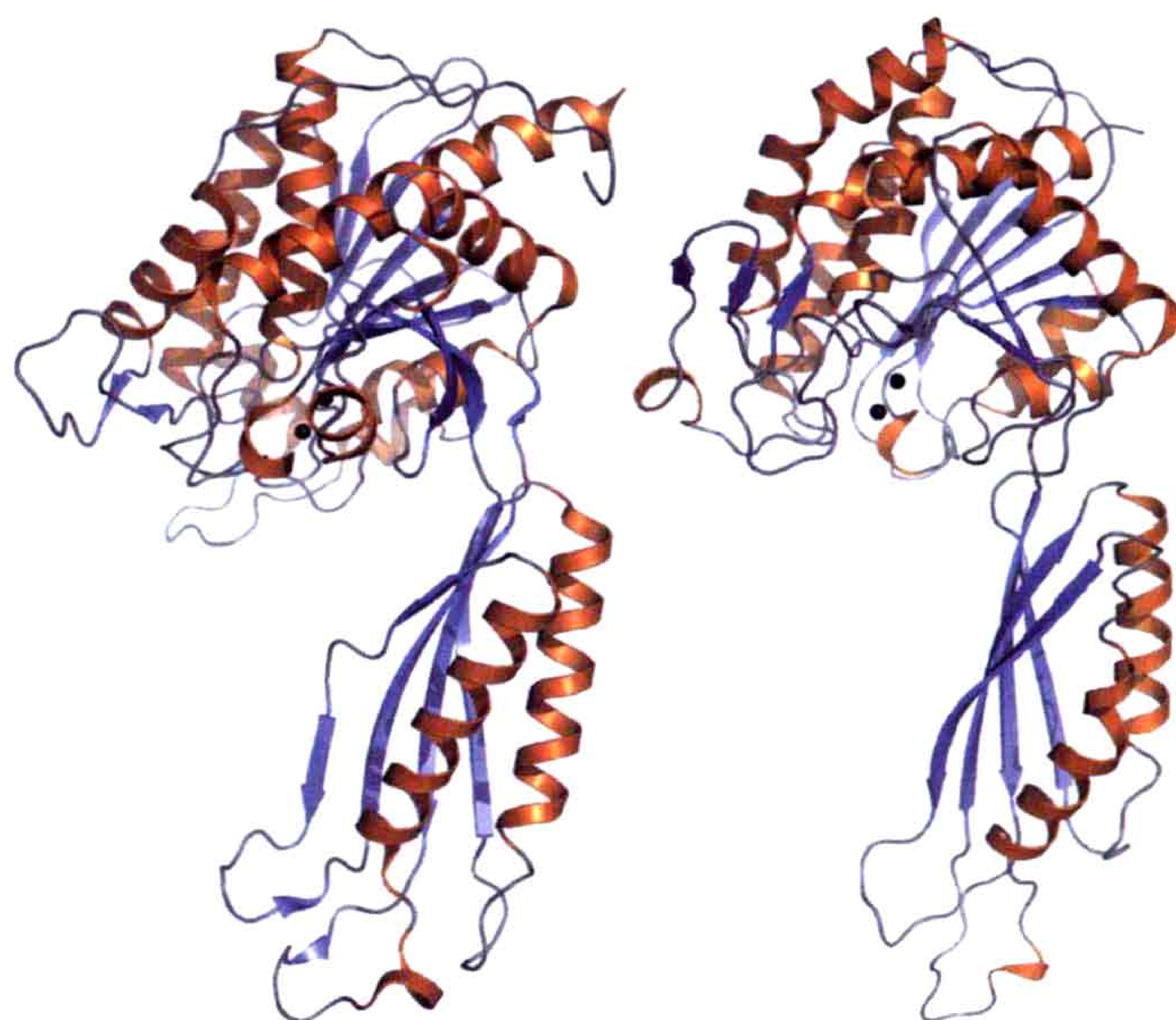


图 16.9　酵母的 BSase（左，PDB: 1R3N）参与催化嘧啶环降解的第三步；细菌的外肽酶（右，PDB: 1CQ2）参与肽键的切割，它们的结构非常相似。

BSase 的例子表明，在解释蛋白质家族的进化历史时，结构生物学展示了强大的力量。甚至在具体的动力学、底物特异性和结构 - 功能开始研究之前，酵母 BSase 的结构加上生物信息学的方法已经提供了一个推测所有 BSase 进化历程的坚实的基础。在酵母 BSase 和一些细菌外肽酶上所观察到的相同的折叠方式暗示这两个家族拥有相同的起源。有趣的是，酵母的 BSase 仍可被二肽和三肽所抑制。很明显，这个祖先可接受从蛋白质中的肽键到嘌呤和嘧啶中 N-C 键的一系列 N-C 键作为底物。

16.4　蛋白质物理化学特性的起源：球蛋白

球蛋白家族可能是从进化角度上研究最多的蛋白质家族。据目前所知，这些蛋白质的功能是结合和运输氧分子。氧分子结合在一个血红素基团上，而血红素基团与蛋白质非共价结合。这些蛋白质由 7 或 8 个 α 螺旋组成，蛋白质主要通过疏水相互作用与位于口袋中的血红素基团结合。蛋白质的一个组氨酸侧链与血红素中心的铁原子相结合。在脊椎动物中，存在两类球蛋白分子：肌红蛋白和血红蛋白。肌红蛋白（图 16.10）是存在于肌肉中的一个单体分子，具有储存氧气的功能。血红蛋白是一个四聚体，由两条球蛋白链，即 α 和 β 链组成，每条链两个拷贝。血红蛋白的 4 个亚基中，每个亚基都结合有一个血红素基团。血红蛋白的功能是从肺中把氧气运输到身体的其他部位。血红蛋白和肌红蛋白是用 X 射线晶体学的方法被分析的第一个蛋白质分子，由 Max Perutz、John Kendrew 和他们的同事完成。

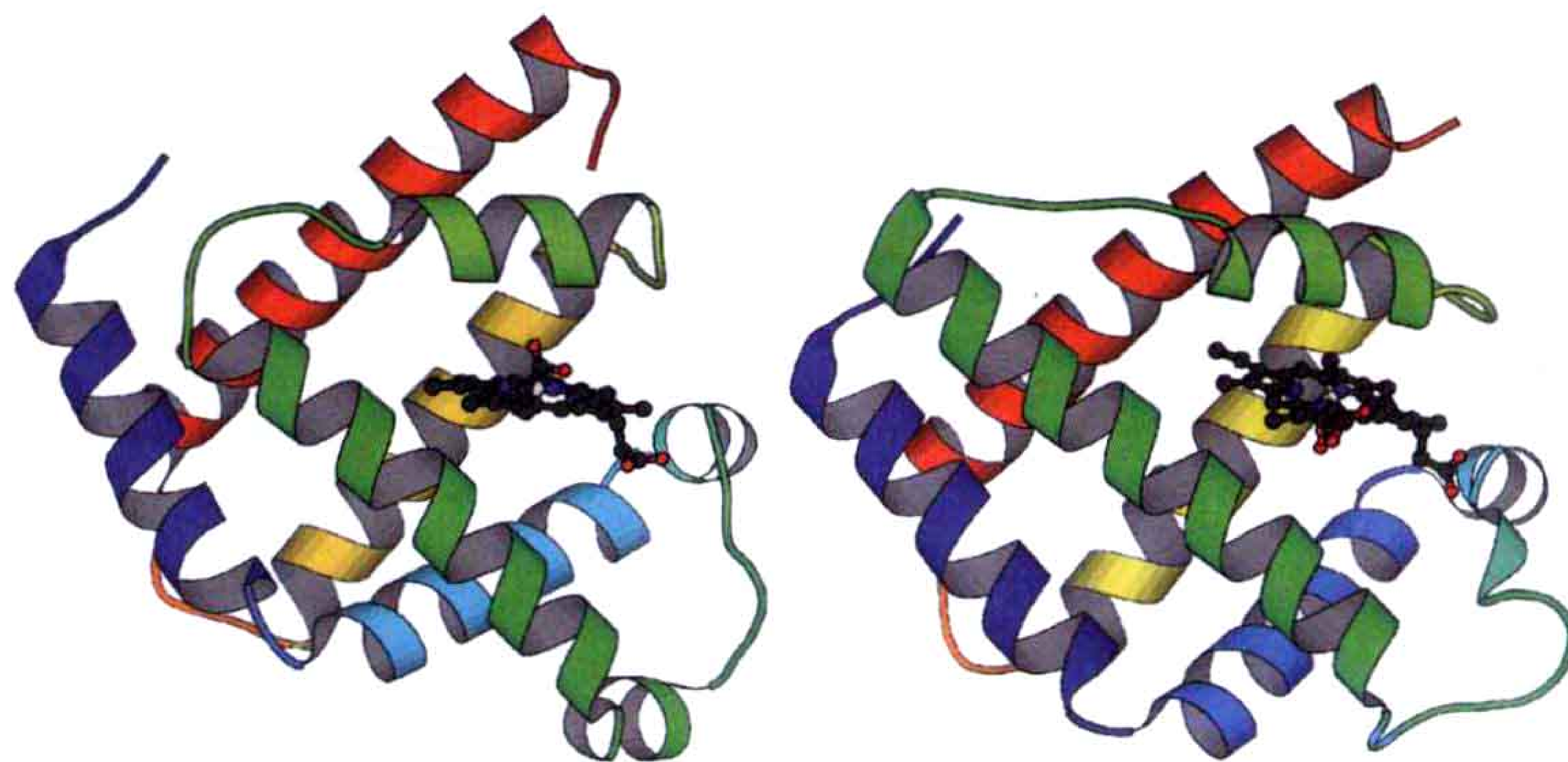


图 16.10　肌红蛋白（抹香鲸）和豆血红蛋白（羽扇豆）（PDB: 5MBN 和 2GDM）。

16.4.1　协同作用的起源

脊椎动物的血红蛋白四聚体具有一些对功能极为重要的特性，对球蛋白序列和结构的深入研究揭示了这些特性发展出来的一条可能的途径。以氧气分压作为变量，氧分子对血红蛋白的结合曲线是“S”形的，在氧分压低时，氧分子结合血红蛋白的能力

相对较弱。血红蛋白氧结合曲线是“S”形的原因是血红蛋白存在两个构象状态：一个是T态，此时所有亚基对氧分子都有一个较低的亲和力；另一个是R态，此时对氧气的亲和力很高。处于T态的亚基与氧分子的结合会引起构象变为R态。当氧分子释放后，血红蛋白分子回到T态。以这种方式，血红蛋白分子可以在氧浓度很高的肺中结合氧气，并把它们有效地运输到身体中氧浓度低的部位。氧气对血红蛋白的亲和性还存在更深入的调控机制，即与pH相关（Bohr效应）。低pH时，蛋白质倾向于形成T态构象，这种构象使氧气在高浓度的乳酸或二氧化碳存在下（这种情形或许是肌肉中的情形）更有效地释放。T和R态之间的平衡同样被各种效应因子所调控。

所有这些效应可归因于血红蛋白四聚体的性质。血红蛋白的协同性是由亚基的相对旋转实现的。在由T态到R态的转变过程中，半个分子，即一对α和β亚基，相对于另一半分子旋转了15°（图16.11）。α和β亚基之间有两类接触：一类是一对亚基内的堆积接触，这对亚基彼此之间并不移动；另一类是亚基间的滑动接触，在构象转换中两对亚基间发生转动。在转换过程中，一些极性相互作用得以形成或受到破坏。在R态时其中一个相互作用是C端的His残基和β链的一个Asp残基之间的相互作用，这个相互作用升高了组氨酸侧链的pK_a值，这与T态的情况相关。这导致了从R态到T态转变的过程中质子的吸收，是Bohr效应的部分基础。随着转变的进行，效应分子如磷酸甘油酸在T态时的结合位点改变了，所以在R态时这种结合并不发生。磷酸甘油酸因此稳定了T态，这解释了氧亲和力对这些分子的依赖效应。

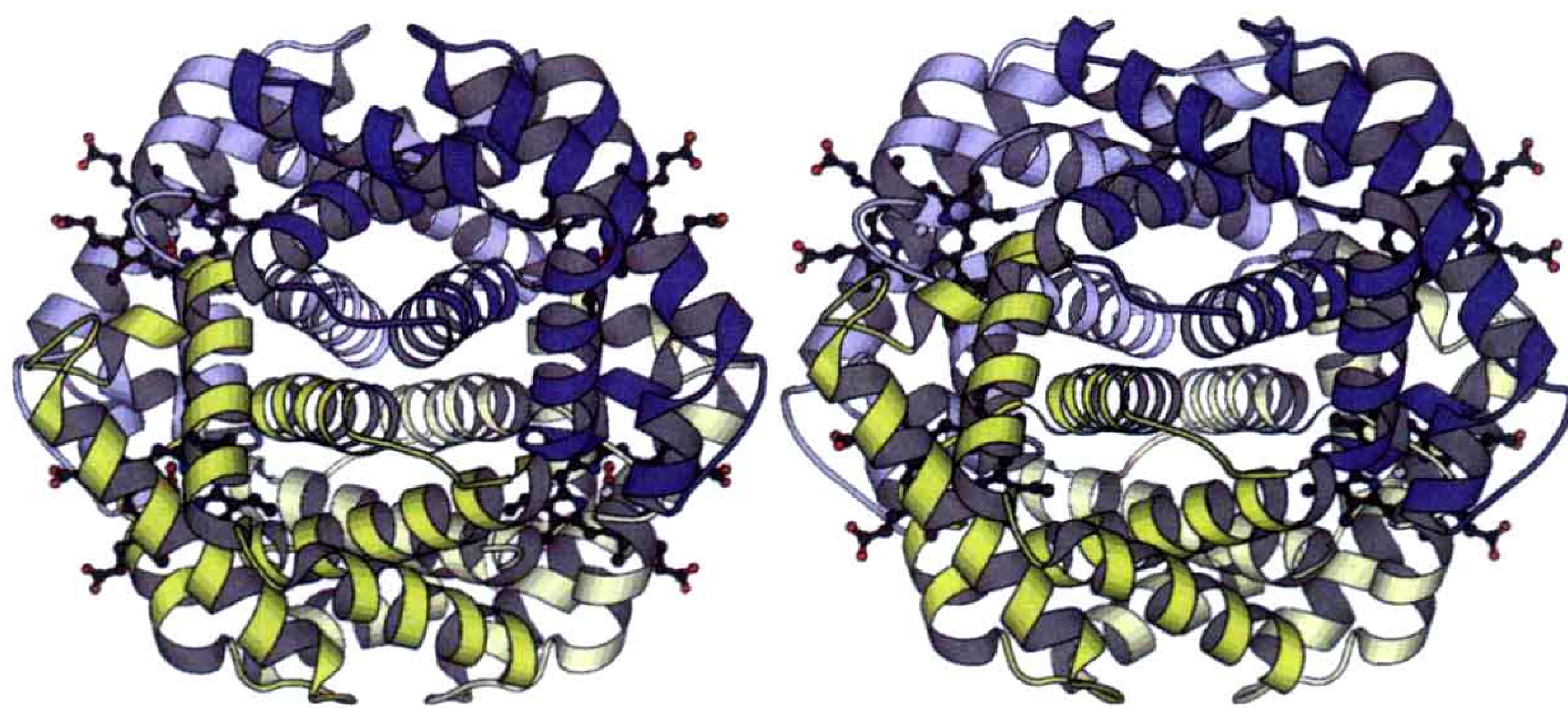


图16.11　血红蛋白中亚基的旋转。图中所示的是一个血红蛋白四聚体的氧结合形式（左，PDB: 1HHO）和脱氧形式（右，PDB: 3HHB）。α亚基为黄色和浅黄色；β亚基为蓝色和浅蓝色。

构象的改变由氧气与血红素的结合引发。当氧气与血红素结合时，会诱导血红素一个很小的构象改变。这将引起亚基产生一个形变力，这个形变力在构象由T态转换为R态时才得以释放。血红蛋白有一种机制使这种构象转换不会轻易发生（图16.12）。在T态时，β亚基的一个组氨酸侧链（HisB97）插入α亚基的一个脯氨酸和ThrA41的两侧链之间；但在R态时，这个组氨酸插入ThrA41和另一个苏氨酸之间。从T态向R态转换的过程中，这个组氨酸必须越过ThrA41，而空间位阻提供了一个屏障使这两个非连续的平衡状态得以清楚的区分。HisB97和ThrA41这两个可滑动的残基在另一对亚基之间的接触位点上对应的是两个精氨酸残基，即ArgA92和ArgB40，它们的体

积非常大因而不能跨过对方滑动。因此，这个亚基间的接触位点只能作为一个固定位点而不能是一个滑动位点。

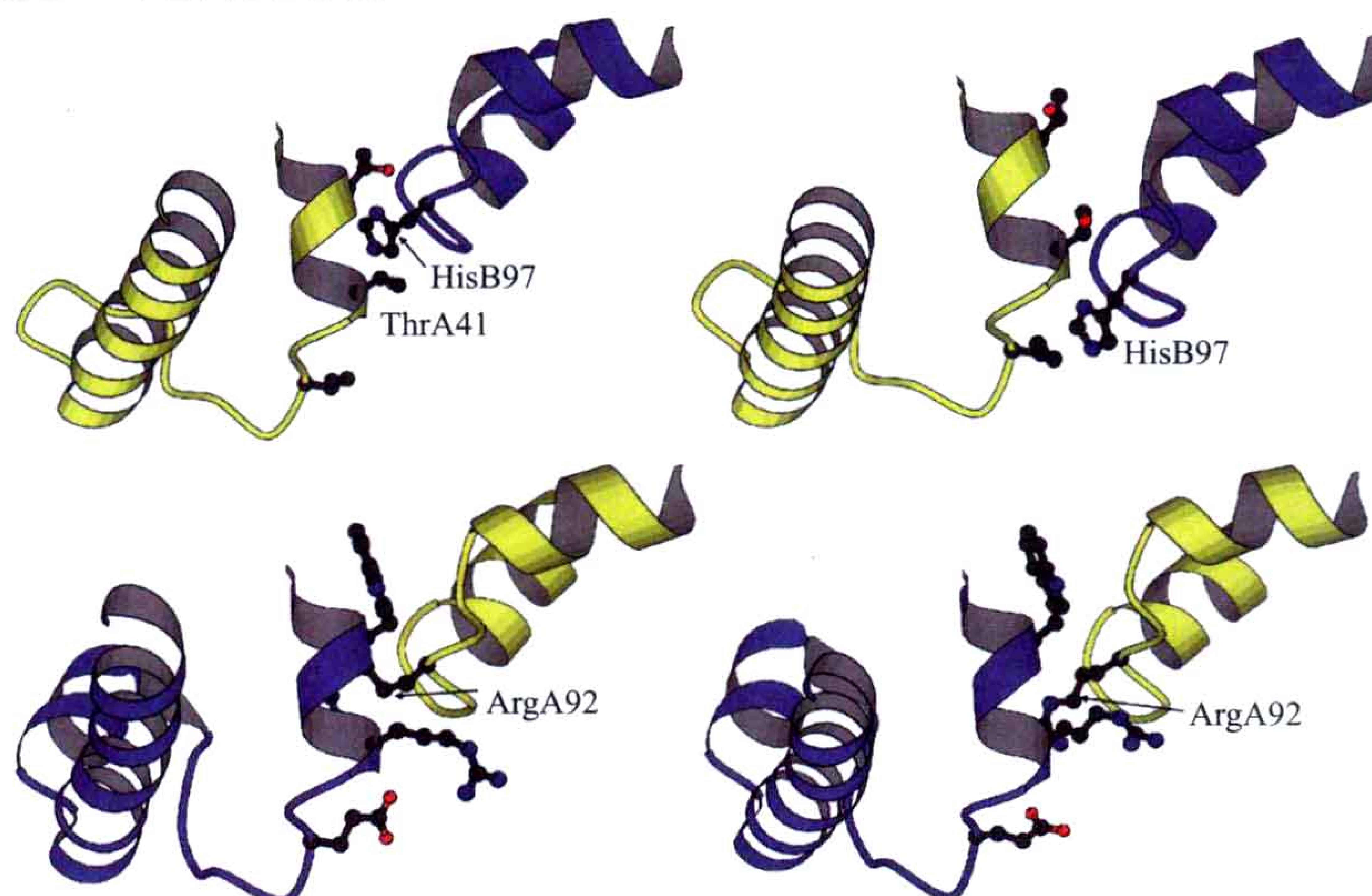


图 16.12　血红蛋白中的滑动接触和固定接触。上：滑动接触。HisB97 必须跨越 ThrA41 才能从氧合形式（左）转变为脱氧形式（右）。下：固定接触。在相应的固定接触中，ArgA92 不能跨越 ArgB40。

具有协同结合氧气能力的血红蛋白四聚体的进化过程包括一个前体基因发生加倍复制以形成 α 和 β 链，以及这些分子发生改变以引入亚基 - 亚基互作界面，还有上述控制两种状态之间的切换机制的产生过程。效应分子结合位点的产生和造成 Bohr 效应的极性互作的产生同样是血红蛋白进化过程中非常重要的变异。

球蛋白分子的序列比对暗示了现今球蛋白链的进化历史。以人类（和其他灵长类）为例，其同样也存在一些除 α 和 β 链之外的其他有功能的球蛋白链基因。它们是在胎儿血红蛋白中代替 β 链的 γ 链、在胚胎血红蛋白中代替 α 和 β 链的 ε 和 ζ 链及在一小部分成年血红蛋白中代替 β 链的 δ 链。

几乎所有的脊椎动物都有一种肌红蛋白和一种四聚体形式的血红蛋白。但一个例外，即七鳃鳗，它有一种具有两个相同亚基的二聚体形式的血红蛋白。当与氧气结合时，二聚体形式的七鳃鳗血红蛋白会发生解离。这种解离包括类似于血红蛋白别构调控转换过程中的构象改变，暗示七鳃鳗血红蛋白可能是四聚体血红蛋白分子的前体。在这个前体中，已经发生了允许形成二聚体的变异。这些表面可以逐渐形成血红蛋白四聚体中的滑动接触面。下一步可能是逐渐形成一个亚基间的更疏水的接触面，相应于血红蛋白中的堆积接触面，使四聚体得以形成。这种改变的一个例证是第 114 位的 His 被 Leu 所替换。最后一步可能是通过侧链的变异产生不甚相同的 α 和 β 链，从而逐渐形成协同作用。现今人类 β 链的一种全 β 四聚体具有一种类似于 R 态形式的亚基排布，并在氧气结合时不发生构象转换。原始的四聚体可能类似于一个全 β 四聚体。α 和 β 链之间最大的不同是 α 亚基中 D 螺旋的缺失。四聚体中两个亚基 D 螺旋的缺失可以使参与转换机制的残基发生紧密的相互作用。

16.4.2 毛蚶（*Scapharca*）血红蛋白：平行进化

利用不同的分子机制实现相同的特性，这个过程被称为平行进化。血红蛋白中引入协同作用的一个完全不同的方式在不等壳毛蚶（*Scapharca inaequivalvis*）中被发现。这个生物有两种协同的血红蛋白，其中一种是同源二聚体。这个二聚体血红蛋白没有脊椎动物血红蛋白中存在的配体效应。这个二聚体由两个球蛋白型的亚基构成，但它是利用脊椎动物血红蛋白中暴露于溶剂中的部分结构结合在一起的，而不是四聚体血红蛋白中形成亚基互作面的部分。血红蛋白结合第一个氧分子后，对氧气亲和力的改变不是由亚基旋转造成。其协同机制是因为处于血红素口袋中的一个苯丙氨酸侧链被挤出，从而血红素结合氧气的能力更强；同时与氧分子的结合影响了二聚体间的相互作用，使另一个亚基中相应的侧链被挤出（图 16.13）。

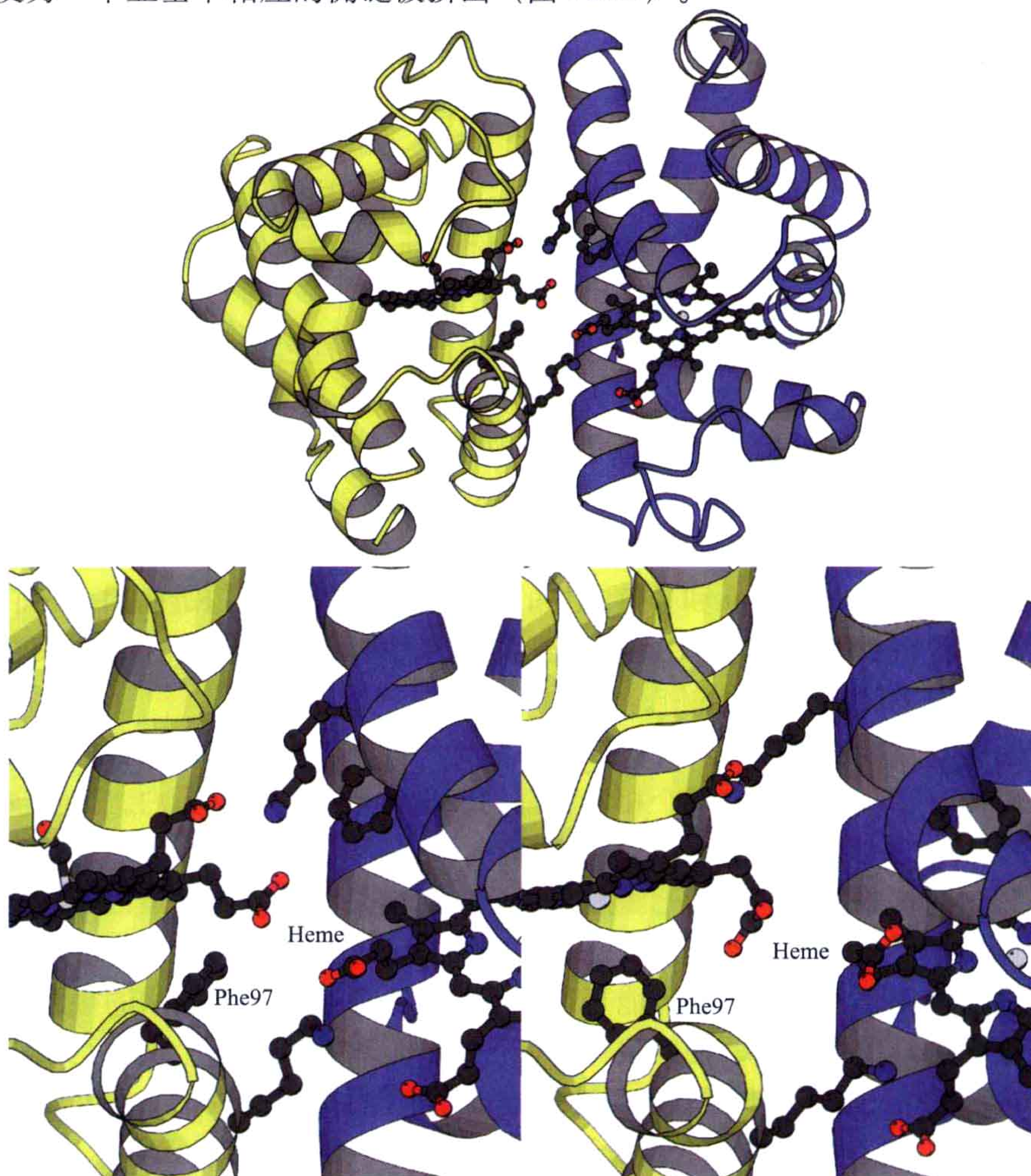


图 16.13 毛蚶二聚体血红蛋白的协同作用。亚基间通过其他面而非脊椎动物血红蛋白的互作面互作，协同作用通过 Phe97 侧链的运动实现。二聚体（上），以及氧合形式（左下）和脱氧形式（右下）的互作面细节图（PDB: 3SDH 和 4SDH）。

16.4.3　血红蛋白中微小的适应性改变

对许多不同动物血红蛋白分子的研究已经表明这些分子对不同条件有适应性改变。在某些情况下，生活在极端海拔环境下的动物具有对氧分子高亲和力的血红蛋白分子，来弥补低的氧气分压。例如，两种鸟类，即喜马拉雅斑头雁和安第斯雁，与它们低海拔下最近的近亲相比，都显示出对氧分子的高亲和力。一项序列差异性分析显示，它们都存在一个替换，这个替换消除了T态下α链Pro142与β链Leu65之间的非极性接触。在斑头雁血红蛋白中，存在一个Pro142Ala的替换；而在安第斯雁血红蛋白中，存在一个Leu65Ser的替换。这些替换将降低T态的稳定性，从而使蛋白质更倾向于形成R态，这将使蛋白质对氧分子的亲和力增加。

这些和其他微小的适应性改变的例证使我们提出这样的问题，即相关物种血红蛋白之间有多少现存的替换可归因于微小的适应性改变，而又有多少可或多或少归因于中性的随机突变呢？在血红蛋白中，大部分观察到的功能性改变可以通过一个或极少数几个替换来解释，而剩下的替换看起来是中性的。然而，这一假设还需通过定点突变的方法来严格检测。一个单一位点的改变就可使分子适应一个新的环境是可能的。例如，使一种动物在高海拔环境下生存，而其他对功能没有太多有利影响的改变也随着时间的推移被引入。

16.4.4　起源

球蛋白，不论是肌红蛋白还是血红蛋白，在所有脊椎动物中都有发现。球蛋白也在非脊椎动物中被发现，如软体动物、海洋蠕虫和昆虫幼虫。一种球蛋白分子，即发现于豆类根瘤中的豆血红蛋白也在很久之前就为人们所了解（图16.10）。最近，球蛋白也在一些没有根瘤的植物、细菌以及酵母中被发现。这些球蛋白中许多是单体，但也有多聚体蛋白和蛋白质链上有多个球蛋白结构域的例子，如卤虫（*Artemia salina*）具有一种包含9个球蛋白结构域的蛋白质链。酵母和大肠杆菌的血红蛋白是一个黄素血红蛋白结构域，这个蛋白质的其他结构域是一个与NADH还原酶相关的黄素蛋白。

一些具代表性的非脊椎动物血红蛋白的三维结构已经被解析。它们的序列相似性非常有限——只有两个残基在所有球蛋白中是保守的，它们是结合在血红素基团中铁原子上的组氨酸残基和一个与血红素基团相对排列的苯丙氨酸残基。但这些蛋白质仍然有保守的特征说明它们是相互关联的。这使人们对这个分子的起源提出这样的问题，即球蛋白看似在祖先分化出现今的细菌和真核生物之前就已经存在了。

所有现存的蛋白质的一个共同的功能看似是先结合血红素随后再结合氧分子。在许多例子中，非脊椎动物球蛋白的确切功能并不清楚。双翅目摇蚊属（*Chironomus*）幼虫的无脊椎动物血红蛋白可能用于给这个生物体的呼吸链提供氧气。豆血红蛋白主要的作用可能是把氧气从根瘤的固氮机器中移除。

球蛋白的进化历史可以从序列和结构的比对中追溯。序列的改变及其对功能的影响在早期事件中当然无从得知。然而，在脊椎动物血红蛋白的例子中，我们已经详细了解了它们的功能性质和很多的序列信息，这使我们可以探讨当四聚体的血红蛋白链产生时发生了什么。

延伸阅读

原始文献

Knecht W, Sandrini MP, Johansson K, et al. (2002) A few amino acid substitutions can convert deoxyribonucleoside kinase specificity from pyrimidines to purines. *Embo J* **21**: 1873-1880.

Lohkamp B, Andersen B, Piskur J, Dobritzsch D. (2006) The crystal structures of dihydropyrimidinases reaffirm the close relationship between cyclic amidohydrolases and explain their substrate specificity. *J Biol Chem* **281**: 13762-13776.

Lundgren S, Gojkovic Z, Piskur J, Dobritzsch D. (2003) Yeast beta-alanine synthase shares a structural scaffold and origin with dizinc-dependent exopeptidases. *J Biol Chem* **278**: 51851-51862.

Royer WE, Jr. (1994) High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. *J Mol Biol* **235**: 657-681.

综述文章

Dickerson RE, Geis I. (1983) Hemoglobin: Structure, Function, Evolution and Pathology. Benjamin/Cummings, Menlo Park, California.

Sandrini MP, Piskur J. (2005) Deoxyribonucleoside kinases: Two enzyme families catalyze the same reaction. *Trends Biochem Sci* **30**: 225-228.

（王子曦 译　苏晓东 校）

附录 A

大分子的化学键及能量关系

A.1 共价键

两个原子之间形成化学键，是由于原子核和电子的重新排布从而得到比单个原子更低的能级。如果低能级的产生是一个或多个电子完全从一个原子转移到另一个原子从而形成带电的离子，那么它们相互间通过静电吸力（库仑力）来形成化合物，就形成了离子键，如 NaCl 的晶体。如果低能级的产生是通过共享电子对来实现的，那么原子之间形成共价键，由共价键相连的分子是不连续的，如 H_2 和 NH_3。对生物大分子来说，共价键是最常见的，氢键也非常普遍存在（见下文），而离子键只是有时会出现在许多蛋白质表面的盐桥中。当共价键形成时，原子间共享电子对，以便得到惰性气体结构（Lewis 的八电子规则）。此外，还有第三种类型的化学键，被称为金属键，它是通过大量的阳离子和众多的电子相互堆积形成的。当原子的最外层电子（价电子）转移时产生的能量变化对于化学键的生成至关重要，也就是说，原子的电子结构变化在成键过程中起到关键的作用。要在理论上描述化学键，就需要利用分子的量子理论和计算。表 A.1 中给出了一些典型的键长。

表 A.1 一些平均键长

键	平均键长 /Å
C—H	1.09
C—C	1.54
C = C	1.34
C—C（苯）	1.39
C ≡ C	1.20
C—O	1.43
C = O	1.12
O—H	0.96
N—H	1.01
N—O	1.40
N = O	1.20

对于键的强度，人们可以通过键能来评估，有时这也被称为键的解离能，就是研究将 1mol 的特定化学键打开所需要的能量。键能通常是在 160 ～ 1100 kJ/mol 之间变化，这取决于原子间的成键个数，如 N_2 具有 942kJ/mol 的三键键能，而 F_2 具有 155kJ/

mol 的键能。在特定体系中，键能具有随原子数的变化而变化的趋势。在同族元素中，键能通常随着原子数的增加而减弱，如 HF > HCl > HBr > HI。如同键长一样，相同原子间的键能在不同化合物中是比较一致的（变化在 10% 以内）。这样便可以通过测量一系列的化合物的键能总结出其平均键能。应该指出当碳碳单键变为碳碳三键时，平均键能从 345kJ/mol 增长为 809kJ/mol。

二硫键

蛋白质的一个特性就是在两个半胱氨酸的硫原子间可以形成共价的二硫键。在还原条件下，二硫键被打开并且硫原子被质子化。对有些蛋白质来说，这种相互作用的氧化还原特性是非常重要的。二硫键中两个硫原子之间的距离是 2Å。为了增加蛋白质的稳定性或是探测其灵活程度，许多二硫键通过突变的方法被引入蛋白质内。当引入这些二硫键时，有必要遵循产生二硫键的自然法则。在二硫键中，两个半胱氨酸的 β 位碳原子 (C_β) 在俯视图中呈 90º 夹角（图 A.1）。

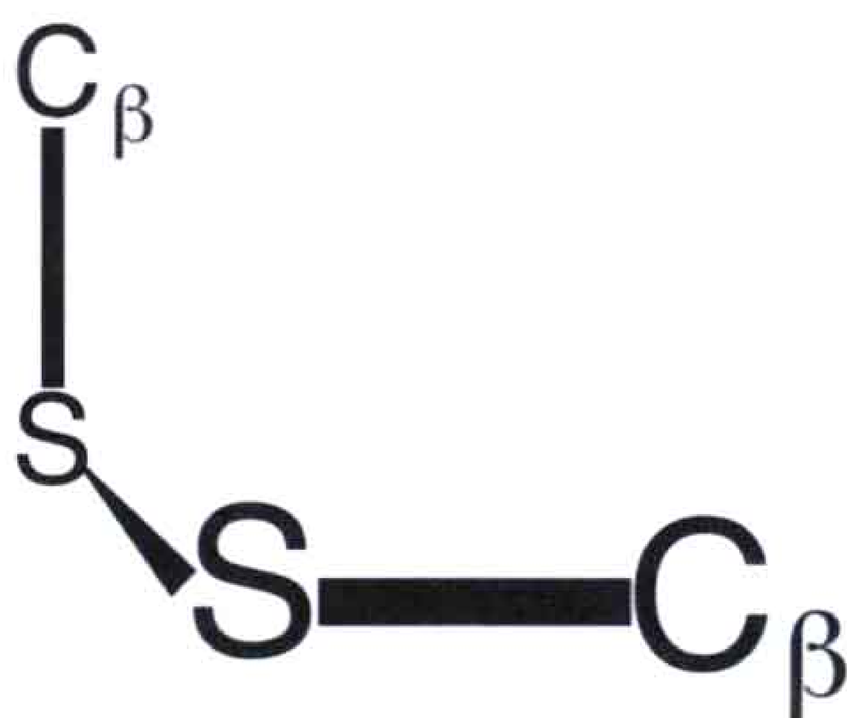


图 A.1　二硫键在 S—S 方向的俯视图中，两个 S—C_β 键在优势构型中呈 90º 夹角。

A.2　电荷相互作用

在生物体系里，静电相互吸引是非常重要但又十分抽象的。在此附录中，我们无法概述整个静电理论，只能略提一二。对于球状蛋白或者脂质膜来说，其表面的带电基团往往是被溶剂化的，也就是说被其反向电荷包围着。这些带电基团包括蛋白质的赖氨酸和精氨酸残基，或脂质膜上的磷酸和乙醇胺及丝氨酸基团。在蛋白质的核心部位或是脂质双层膜中包埋一个孤立的带电基团不仅将导致其溶剂化能量的丢失，而且会以巨大的静电损失为代价。带异种电荷的离子之间是通过库伦力来相互吸引的，也就是正比于 $e^2/\varepsilon r$，在这里 e 是单位电荷，r 是电荷被当做点电荷时两个电荷之间的距离，ε 是电荷所处介质的介电常数。在水溶液中，介电常数 ε 约等于 80。而在蛋白质内部或是脂质双层膜中，介电常数 ε 则低得多（只有 2 ~ 20）。这样，如果将一个电荷和

同它匹配的异种电荷一起包埋在蛋白内部较近的距离内，就会产生巨大的能量优势。这就是盐桥产生的原因，并且解释了为什么需要跨膜的通道来传输离子的原因。

当考虑到与蛋白质或膜表面电荷相互作用并结合的离子时，静电理论变得十分复杂。同样的情况也存在于当我们考虑两个带相同或不同电荷的膜之间的作用时，或一个膜是中性而另一个带电荷时。在这些情况下，需要讨论渗透压和熵的因素，本书将不做介绍。

A.3　极性相互作用

A.3.1　氢键

氢键在生物体中十分重要。它们的形成有利于化学基团之间的稳定和取向。氢键是在质子与两个邻近的具有孤对电子的电负性原子（供体与受体）间形成的。这也表明质子是同时结合于两个电负性原子的。氢键供体与受体的取向和距离是多样的。氢键的强度主要与其长度和线性相关。通常情况下，氢键的键长约为 2.8Å，生成焓约是 20kJ/mol。而氢键的键角对于成键的强弱至关重要。供体、质子和受体之间的键角一般近似于 180°。如果偏离这个角度，则会降低键的强度。受体原子的键角（如当氢键是通过羰基上的氧来形成时，C—O—H 之间的夹角）通常是 120°，这对应于氧原子上孤对电子对的取向。这个角度的偏离不似前者那么重要。例如，在通过氢键来稳定的蛋白质二级结构中，这个角度往往接近 180°，这可能是由空间受限造成的。

在大分子中，氢键的供体通过共价键与氢原子相连，而受体具有孤对电子。氢键的供体和受体通常是氮原子和氧原子，也有其他原子，如半胱氨酸中的硫原子。不同类型的静电相互作用可以在 C—H 基团和 π 电子之间产生。下文中将描述一个这样的例子，它是在芳香基团中产生的。

大分子物质中存在大量氢键的供体与受体。在蛋白质或是核酸分子的内部，这些氢键的受体总是会找到与其相应的供体。如果氢键的形成是以较高的能量损失为代价，那么这个氢键将会是不稳定的。癌症的抑制因子 p53 就是一个高度不稳定的蛋白质（见 7.6.5 节）。它的不稳定性可能是由于在蛋白质内部的极性基团不能够正确地形成氢键而造成的。

A.3.2　低能垒氢键（LBHB）

氢键的势能图中通常有两个势井。这两个势井可以是不同深度的［图 A.2(a)］。这取决于氢键中的质子是与供体还是与受体相连就是可变的。然而，两个势井也可以是

同等深度的［图 A.2(b)］，或是两个势井之间的能垒很低甚至不存在［低能垒氢键，图 A.2(c)］。氢键的供体与受体 pK_a 值之间的关系对氢键相互作用强度的影响非常关键。二者之间 pK_a 值越接近，则其产生的氢键也就越强。在这种情况下，可以得到键长为 2.3Å 的氢键，它在气相中的键能是 100 ~ 125kJ/mol，大约为其在水中键能的一半。这种低能垒氢键 (LBHB) 具有相当可观的共价贡献。

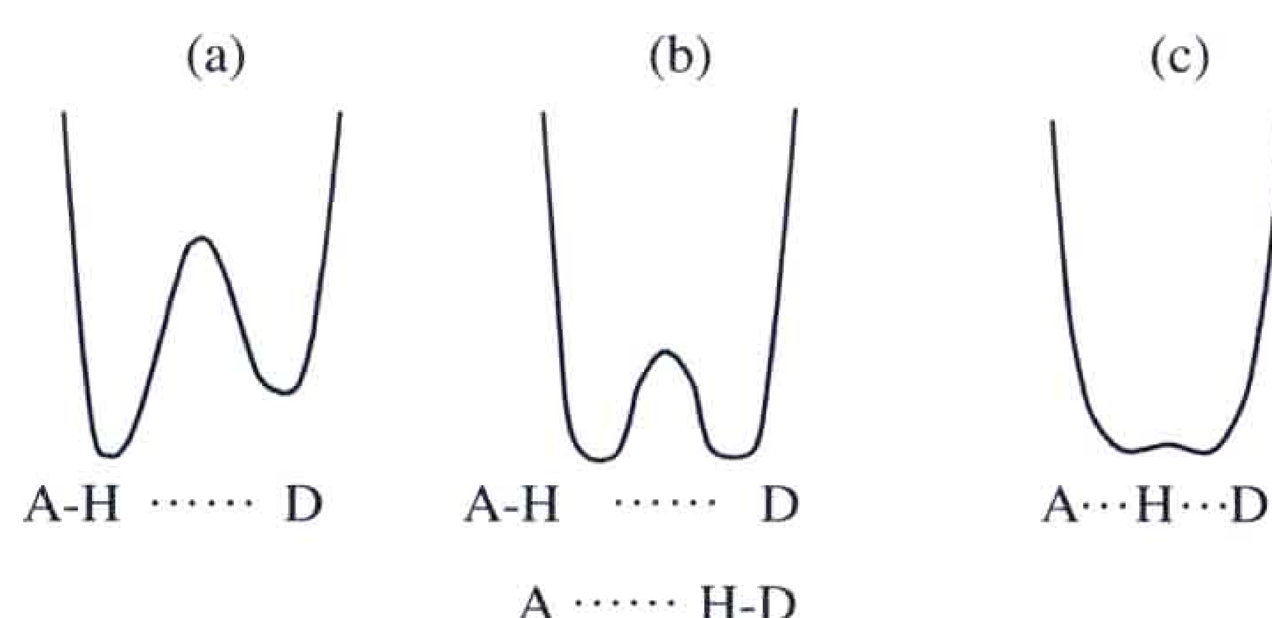


图 A.2　氢键中质子与供体和受体相互作用的势能函数。(a) 具有两个不同势井深度的氢键，质子通常是结合在供体而非受体之上；(b) 具有两个相同势井深度的氢键，质子可以与供体结合，也可以与受体结合；(c) 低能垒氢键 (LBHB)，质子在供体与受体之间平均分配。

通过 X 射线晶体衍射来确定氢键是有难度的，往往只能在高分辨率的晶体中观察到。供体与受体原子之间的距离可以被测定，但其精确度则取决于分辨率的大小。因此，距离较短的氢键 (2.5Å 以内) 可以在高分辨率的结构中被确认。短距离氢键也可以通过 NMR 来寻找，因为它们可以在 NMR 图谱中显示明显的化学位移。红外图谱也可以揭示不寻常的氢键。研究氢键最好的实验手段可能要数中子衍射了。这一方法可以用来分辨氢和氘。在低能垒氢键 (LBHB) 中，氢原子可能没法明确地定位：它可以在供体与受体原子之间任意存在。也可以用计算化学的办法来指认低能垒氢键 (LBHB)。此外，对这些较短的氢键来说，氢 / 氘的交换是十分受限的。

酶的活性位点往往是不与水接触的，因此氢键比较强。对于一般的氢键来说，要使质子接近其受体需要克服一个显著的能垒 (图 A.3)，这将导致质子的迁移速率很低。在低能垒氢键 (LBHB) 中，质子是在供体与受体之间平均分布的，因此有很高的迁移速率。

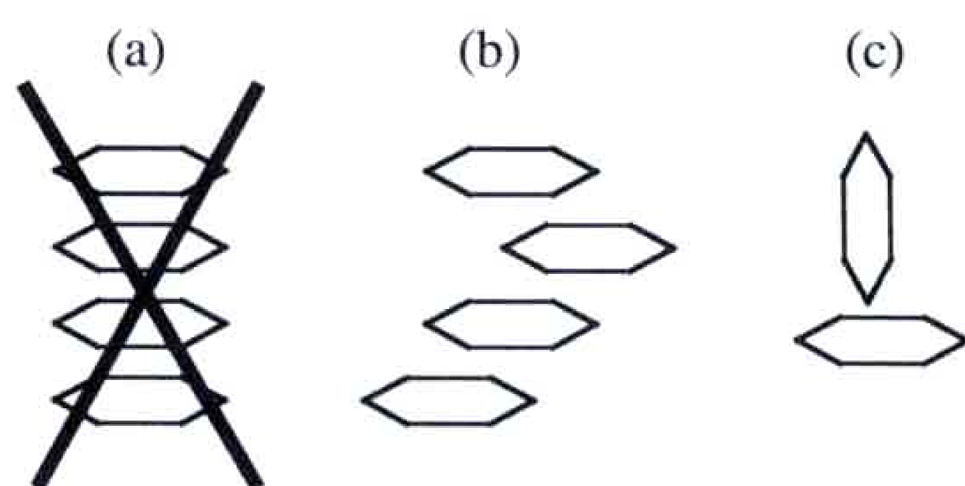


图 A.3　芳香族相互作用。(a) 将芳香基团直接堆积到彼此顶部是不可取的；(b) 向侧面位移的堆积往往可以提供理想的相互作用；(c) 芳香基团也比较喜欢以正交的方式相互作用。

在好几种酶中都发现有低能垒氢键 (LBHB) 的存在，其中包括酸性的蛋白酶。HIV-1 蛋白水解酶是一个同源二聚体，其中每个亚基上的 Asp25 残基组成了一个催化二体。不同的短氢键已经在活性位点的残基间被发现，这取决于配体和酶的状态。同样，在丝氨酸蛋白酶中，氢键存在于由 Asp 和 His 组成的催化三体中，在 NMR 图谱中显著的化学位移说明这个氢键很短。在酶的某些状态下，这个氢键可以被认为是低能垒氢键 (LBHB)。

A.3.3 芳香族相互作用

核酸（如碱基）或蛋白质（如 Trp、Tyr、Phe 和 His 的残基侧链）可以提供特殊的相互作用。位于芳香环两侧的 π 电子云可以提供部分负电荷，而芳香基团的周边则具有相应的部分正电荷。这样的部分电荷主要通过两种方式在芳香基团之间起到稳定作用。一种方式是芳香基团在彼此顶端堆积。一个理想的芳香基团的堆积并不是如图 A.3(a) 所示的堆积，而是将带正电荷的芳香环边缘迁移到能够与相邻芳香环上带负电荷的 π 电子云相互作用的位置[图 A.3(b)]。另一种方式则是两个相互作用的芳香环相互正交[图 A.3(c)] 。这样，一个环的边缘又一次朝向另一个环的表面。

糖基也可以通过糖环一侧羟基的吸电子效应在糖基上没有羟基的表面产生部分正电荷。这个带部分正电荷的表面倾向于通过电荷间的相互作用及范德华力与芳香族基团相互作用 (图 A.4)。

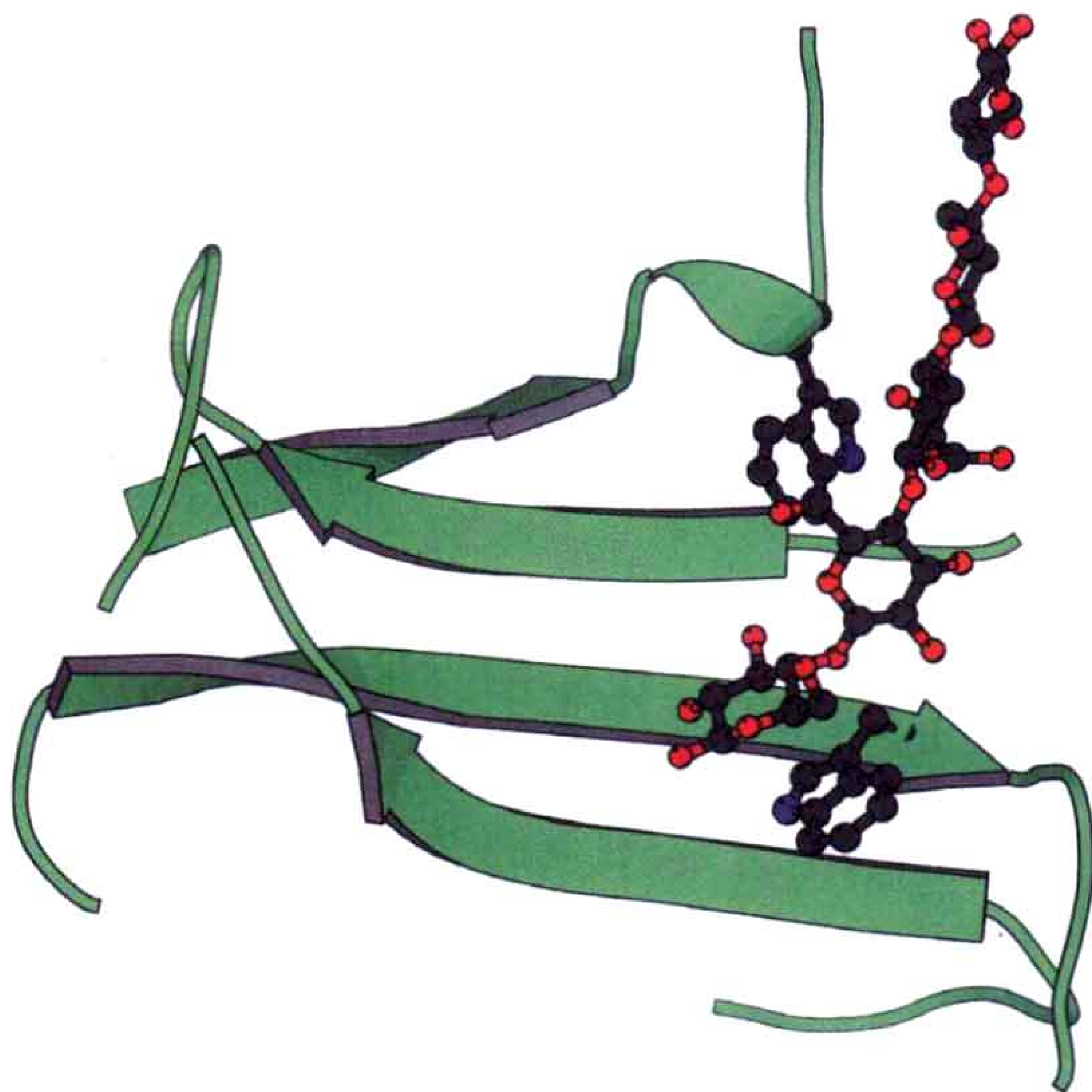


图 A.4 芳香基团与碳水化合物的相互作用：位于纤维素酶 (cellobiohydrolase，CBH1) 通道表面的两个色氨酸残基可以和纤维戊糖结合 (PDB:6CEL)。

A.4 非极性相互作用

A.4.1 范德华相互作用

蛋白质的主链是高度极性的，而其许多侧链则是非极性或是疏水的。在可能的程度上，这些侧链间彼此相互作用，而不与极性溶剂水相作用。这就使得疏水性侧链通过范德华相互作用被包埋在蛋白质的内部，原子间的相互距离大约是 3.6Å。生物大分

子中大量存在的范德华相互作用对其稳定性起到了显著的作用。

A.4.2 疏水作用

蛋白质折叠的稳定性是由许多因素决定的，其中最重要的是疏水核心的形成。这一核心的形成是由于非极性侧链在蛋白质或其结构域内的堆积。疏水核心主要不是通过疏水基团相互的吸引来形成的，而是通过其熵的特性导致的疏水效应来实现的。疏水效应归结为蛋白质与其周围的溶剂水分子之间的相互作用。在没有折叠的状态中，蛋白质的疏水侧链是暴露在溶剂中的。而在折叠的状态下，蛋白质中的大部分疏水侧链则与溶剂分子隔离，从而致使水分子的熵增加。疏水核心中的极性基团往往参与氢键或是盐桥的形成。

位于多肽主链的极性基团通常以有序的模式在核心部位形成氢键，如β折叠和α螺旋。由于将极性基团包埋于疏水核心在能量上是不利的，对于大多数的蛋白质来说，其核心部位中蛋白质骨架上所有的极性基团都参与了α螺旋和β折叠中氢键的形成。

A.5 酶活性位点的标准几何偏离

在旧键断裂或新键生成的过程中，位于酶上的特殊原子间的距离和角度将会有两种以上的状态。不同的状态对应于较大的构象变化。这样的过程必然存在中间状态，其键长会瞬时短于范德华键或者长于标准共价键的距离。同样的，键角也会偏离其常规位置，这是由一些还没有完成的旋转造成的。一个典型的例子是丝氨酸蛋白酶，位于活性位点的丝氨酸是以共价结合的方式与底物相互作用的。目前已经通过一些办法使得反应停留在中间体阶段并且获得了其晶体结构。这些办法包括使用不适宜的pH、底物类似物，以及一些其他的手段来进行反应，从而使得反应不能进行完全。显而易见，在研究过程中需要对标准几何偏离做出合理的解释。简单提一下，丝氨酸蛋白酶、乙醇脱氢酶、儿茶酚 *O*- 甲基转移酶等的反应中间体都已得到了确认。低能垒氢键极有可能出现在这些中间体或是过渡态阶段。

延伸阅读

Baker EN, Hubbard RE. (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* **44**: 97-179.

Bürgi HB, Dunitz JD, Shefter E. (1973) Geometrical reaction coordinates. II.

Nucleophilic addition to a carbonyl group. *J Am Chem Soc* **95**: 5065-5067.

Canadillas JMP, Tidow H, Freund SMV, Rutherford TJ, Ang HC, Fersht AR. (2007) Solution structure of p53 core domain: Structural basis for its instability. *Proc Natl Acad Sci USA* **103**: 2109-2114.

Cleland WW, Frey PA, Gerlt JA. (1998) The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* **273**: 25529-25532.

Coates L, Erskine PT, Mall S, Gill R, Wood SP, Myles DAA, Cooper JB. (2006) X-ray, neutron and NMR studies of the catalytic mechanism of aspartate proteinases. *Eur J Biophys* **35**: 559-566.

Flemming PJ, Rose GD. (2005) Do all backbone polar groups in proteins form hydrogen bonds? *Protein Science* **14**: 1911-1917.

Morozov AV, Kortemme T, Tsemekhman K, Baker D. (2004). Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc Natl Acad Sci USA* **101**: 6946-6951.

Schultz CN, Warshel A. (2004) The low barrier hydrogen bond (LBHB) proposal revisited: The case of the Asp···His pair in serine proteases. *Proteins Struct Funct Bioinform* **55**: 711-723.

Tylska J, Groshowski P, McCammon JA. (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protase. *Protein Science* **13**: 513-528.

数据库链接

Hydrogen and hydration database: http://hhdbol.tokai-sc.jaea.go.jp/HHDB/

附录 B
折叠类型的比较方法

B.1 结构比较的方法

结构比对可以用来比较复杂的结构，从而辨别出肉眼不易看到的相似性。这些相似性有助于揭示未知进化过程中的相互联系及功能的相似性。基于结构叠加的序列比对在阐明相互关系时可以提供非常有用的信息，特别是当序列比对相似性较低时。

结构比对不像序列比对那样直截了当。两两比对是基于这样的假设，即两个结构中的残基是可以相互重叠的。如果蛋白质之间有明显的序列相似性，那么起始的相似残基可以通过序列比对来获得。通过这些残基上 Cα 的坐标，可以计算得到最佳的叠加方式，使得一个分子以刚体的形式与另一个分子叠加。用于叠加的残基可以通过预先比对的办法加以修正，这样，那些相互之间在合适的起始位置的其他残基就被包括进来。经过一系列的叠代，得到的是一个叠加结果，包括两个结构中的部分残基和这些残基距离的均方根 (r.m.s.)。有时需要将一个蛋白质分子分割成数量有限的几个片段（往往

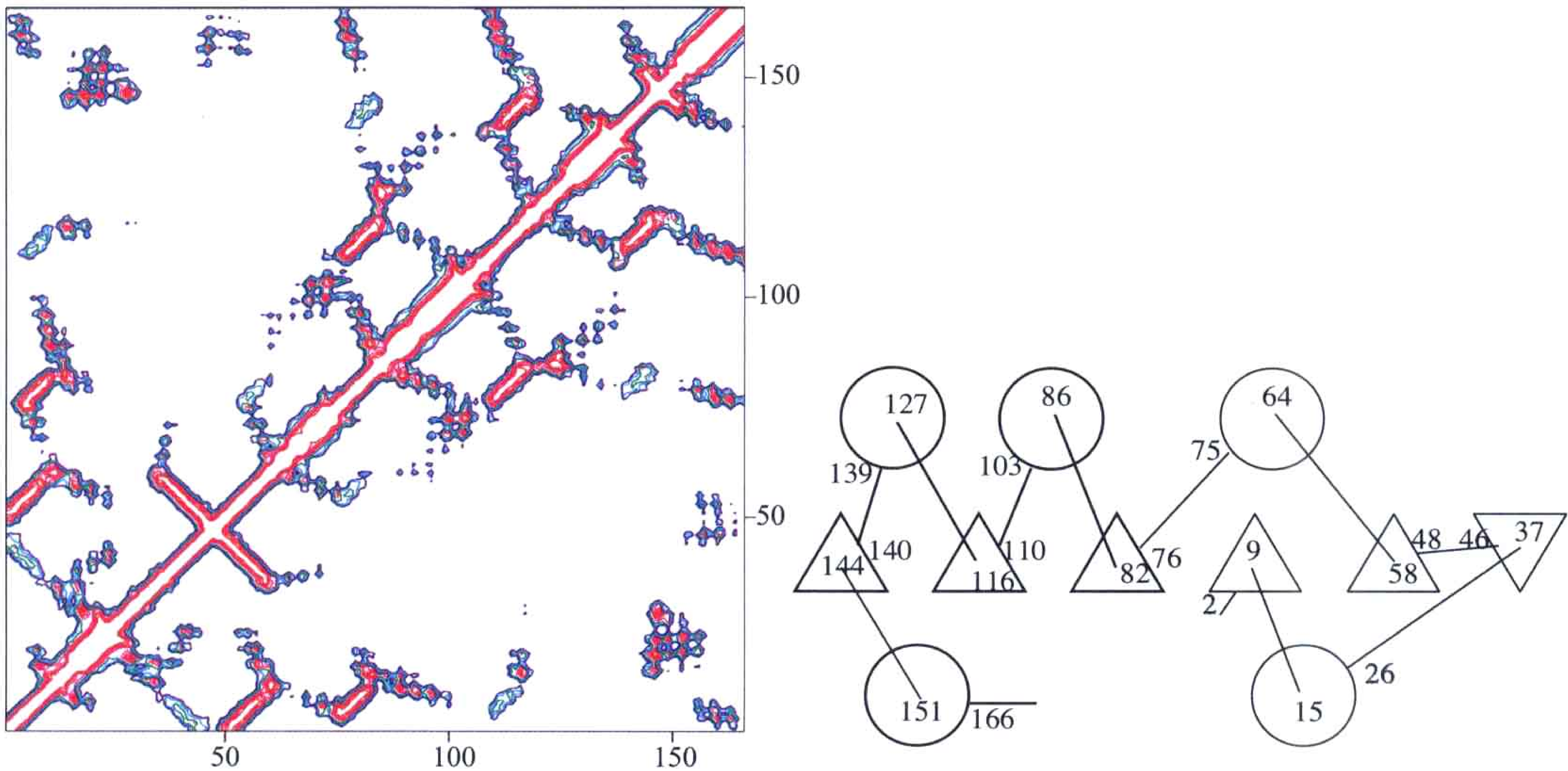


图 B.1　左：Ras 蛋白的距离图谱。等高线表明空间接近的 Cα 碳原子。这个图谱表明蛋白质结构中存在 5 个螺旋结构（沿对角线的厚粗部分），以及 5 个平行折叠和 1 个反平行折叠。Ras 蛋白的结构部分是由 βαβ 单位组成的。右：Ras 蛋白的折叠示意图，表明在每个二级结构元素从起始到终止位点的氨基酸残基的数目（螺旋：圆圈；β 折叠：三角）。距离图谱的特性可以和折叠示意图相互关联。

与它们的结构域一致），然后将这些片段分别叠加以便结构域的旋转。

当不存在明显的起始叠加或是需要将结构与整体蛋白结构数据库比较时，就需要一个更为通用的办法。由于可能在最终比对中存在间隙和插入片段，这就需要一个对蛋白质折叠的简化描述，也就是只比较结构的组成元素，而不是整个分子。距离矩阵就是一个这样的描述，它描述的是每两个残基之间的距离。图 B.1 所描述的是相邻的 Cα 原子。螺旋结构是沿着对角线的厚粗部分的。β 折叠则是那些与对角线垂直（反向平行）或与对角线平行 (平行) 的线。当存在略有距离的接触时，邻近的二级结构元素便显而易见了。很容易看到相似的结构具有相似的距离矩阵。一些叠加的办法正是将结构片段配对并对它们的距离图谱加以比较。在应用最广的 Dali 比对中，两个蛋白质中的片段是以六肽的形式进行配对的。两个配对元素的得分是相似的，可以通过适当的采样数将短片段延长为较长的比对片段。

众多折叠比较的程序都用到一个简便的替代办法，那便是只考虑螺旋与折叠并且寻找那些具有相似取向的结构子集。

B.2　折叠的数据库

B.2.1　蛋白质数据库

蛋白质数据库是蛋白质结构信息的主要来源，它是由 RCSB PDB(美国)、MSD-EBI(英国) 和 PDBj(日本) 组成的国际联盟 wwPDB 支持的。这个数据库存储了由实验测定的蛋白质和核酸分子中原子的坐标。这些信息主要是从由 X 射线晶体衍射及核磁共振波谱所测定的结构中得到的。数据库基本上涵盖了由以上技术得到的大分子化合物的全部结构信息。大多数杂志都要求将要发表结构的坐标提交到 PDB，但作为选择，可以在提交一年后再公布数据。实际数据库的内容包括实验者提交的数据及由 wwPDB 员工所加的注解。另外，对于实验的衍射数据还经常出现在 X 射线衍射的结构中。图 B.2 列举了 PDB 中的条目。

B.2.1.1　实验模型的质量

图 B.2 所列 PDB 中的条目是基于实验数据解释的分子模型。在坐标中有统计误差存在，但遗憾的是在解释中并没有体现出误差来。误差可以分为两类：由实验数据的局限性引入的随机误差和由实验者不同引入的系统误差。在数据被公布之前，这些提交的数据要经过一系列的测试来检测错误及反常特性。这一过程被称为认证，它将对模型作为一个整体或对其单个残基来说的正确性或质量提供重要的信息。认证过程最

重要的结果则包含在条目的开头或通过链接提供。Ramachandran 图谱 (见 2.1 节) 是一个非常有效的分析手段，它是针对蛋白质所有残基的主链构象扭转角度的散射图谱 (图 B.4)。一个好的模型将使这些角度几乎都位于图谱的偏好区域内。如果不是这样，则需要怀疑模型是否正确。

```
HEADER    VIRUS/VIRAL PROTEIN                     05-APR-05   1ZA7
TITLE     THE CRYSTAL STRUCTURE OF SALT STABLE COWPEA CHOLOROTIC
TITLE    2 MOTTLE VIRUS AT 2.7 ANGSTROMS RESOLUTION.
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: COAT PROTEIN;
COMPND   3 CHAIN: A, B, C;
COMPND   4 SYNONYM: CAPSID PROTEIN, CP;
COMPND   5 ENGINEERED: YES;
COMPND   6 MUTATION: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: COWPEA CHLOROTIC MOTTLE VIRUS;
SOURCE   3 ORGANISM_COMMON: VIRUS;
SOURCE   4 GENE: RNA4;
SOURCE   5 EXPRESSION_SYSTEM: VIGNA UNGUICULATA;
SOURCE   6 EXPRESSION_SYSTEM_COMMON: COWPEA;
SOURCE   7 EXPRESSION_SYSTEM_STRAIN: CALIFORNIA BLACKEYE;
SOURCE   8 EXPRESSION_SYSTEM_VECTOR_TYPE: RNA
KEYWDS    MUTANT VIRUS CAPSID STRUCTURE, ICOSAHEDRAL PARTICLE,
KEYWDS   2 STABLIZING MUTATION, STABLE MUTANT, BETA HEXAMER, BETA
KEYWDS   3 BARREL, BROMOVIRUS, POINT MUTATION
EXPDTA    X-RAY DIFFRACTION
AUTHOR    B.BOTHNER,J.A.SPEIR,C.QU,D.A.WILLITS,M.J.YOUNG,J.E.JOHNSON
REVDAT   1   21-MAR-06 1ZA7    0
JRNL        AUTH   J.A.SPEIR,B.BOTHNER,C.QU,D.A.WILLITS,M.J.YOUNG,
JRNL        AUTH 2 J.E.JOHNSON
JRNL        TITL   ENHANCED LOCAL SYMMETRY INTERACTIONS GLOBALLY
JRNL        TITL 2 STABILIZE A MUTANT VIRUS CAPSID THAT MAINTAINS
JRNL        TITL 3 INFECTIVITY AND CAPSID DYNAMICS
JRNL        REF    J. VIROL.                    V.  80  3582 2006
JRNL        REFN   ASTM JOVIAM  US ISSN 0022-
...
ATOM      1  N   GLN A  40     127.326 141.523 188.649  1.00 78.03           N
ATOM      2  CA  GLN A  40     126.941 142.963 188.796  1.00 78.71           C
ATOM      3  C   GLN A  40     126.007 143.163 190.001  1.00 77.96           C
ATOM      4  O   GLN A  40     125.985 142.326 190.932  1.00 79.06           O
ATOM      5  CB  GLN A  40     126.243 143.450 187.516  1.00 78.74           C
ATOM      6  CG  GLN A  40     124.899 142.758 187.236  1.00 79.59           C
ATOM      7  CD  GLN A  40     124.192 143.322 186.009  1.00 79.24           C
ATOM      8  OE1 GLN A  40     124.588 143.058 184.869  1.00 81.35           O
ATOM      9  NE2 GLN A  40     123.138 144.104 186.239  1.00 79.15           N
ATOM     10  N   GLY A  41     125.239 144.262 189.970  1.00 76.76           N
ATOM     11  CA  GLY A  41     124.308 144.563 191.051  1.00 72.98           C
ATOM     12  C   GLY A  41     122.914 144.020 190.777  1.00 69.88           C
ATOM     13  O   GLY A  41     121.981 144.798 190.541  1.00 69.83           O
```

图 B.2　PDB 条目 (1ZA7) 中标题和开头所列原子坐标所示的起始部分。条目中给出了头两个残基 (Gln A40 和 Gly A41) 的坐标。前三列的数字是原子相对于合适坐标系的坐标 (Å)。第四列是占位 (通常状态下是完全占位，用 1.0 表示)。在高分辨率的结构中，一些侧链可以被指认为可互换构象，这样原子的多重坐标就会在相应的占位中被列出。另外，结合的分子可能会有部分占位的情况，这也会在这些数字中被反映出来。接下来的一列数字给出了热力学无序程度的评估 (B- 因子，图 B.3)。

对于 X 射线衍射结构来说，数据的分辨率对所得模型来说，不论是在随机误差还是在错误风险上，都是很好的质量评判标准。分辨率衡量的是实验所得衍射数据的详细程度，数值越小分辨率越高，这就意味着数据可以用来描述越精细的结构信息。如果分辨率高于 2.0Å，所得模型通常被认为是可以信赖的。有些情况下分辨率甚至高于 1.0Å：这种“原子级别分辨率”的模型是非常精确的，包含了氢原子位置和侧链可变构象的详细信息。倘若在结构测定过程中加以特殊处理，分辨率在 3 ～ 4Å 的模型仍然是正确和可用的。表 B.1 给出了在分辨率不断提高时所能评估的分子特性。

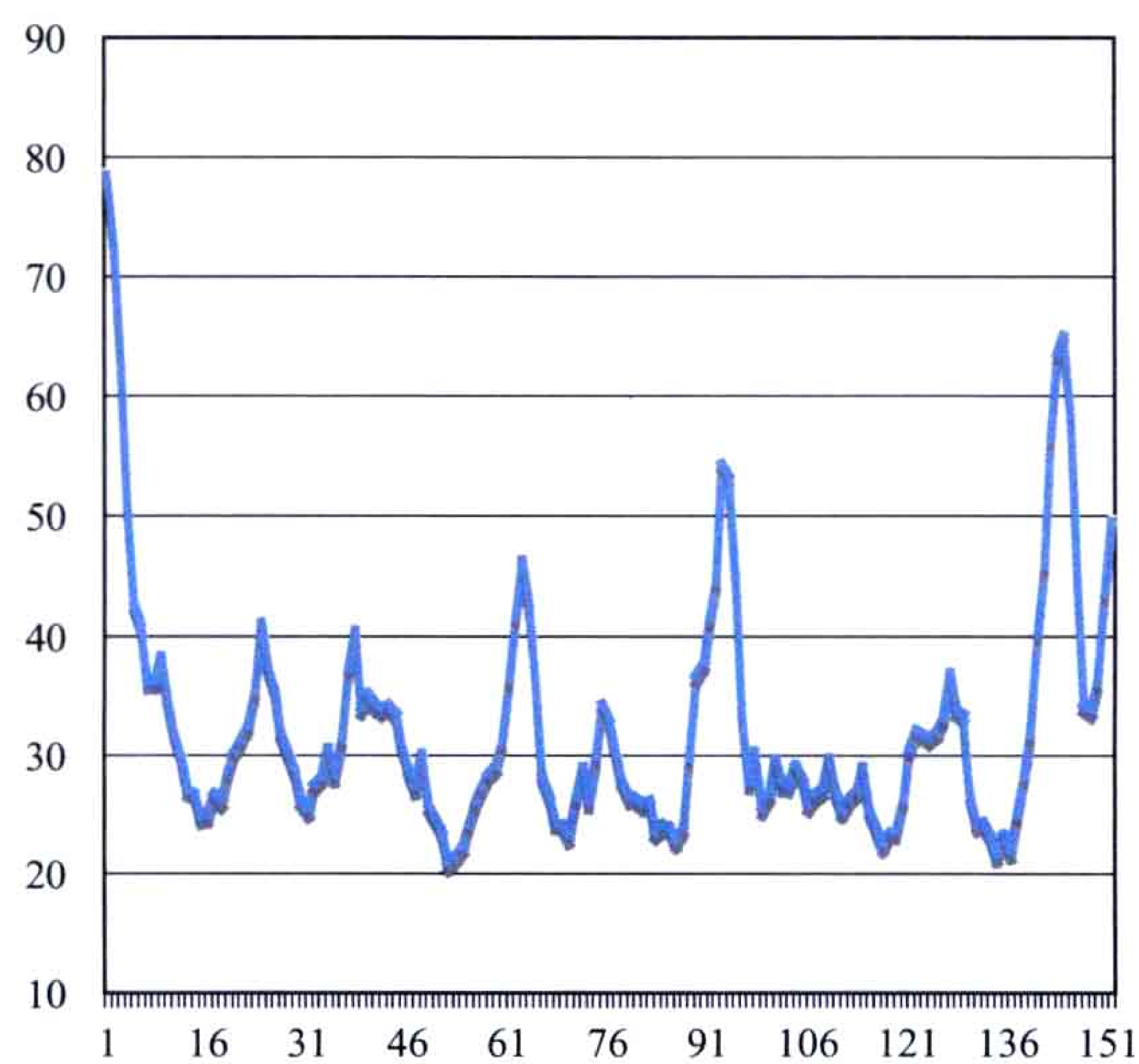


图 B.3　Ras 的 Cα 原子的 B- 因子图谱(条目 1ZA7 中三条等同链中的一条)。在表面环状结构中或在链的终端，B- 因子通常比较高。

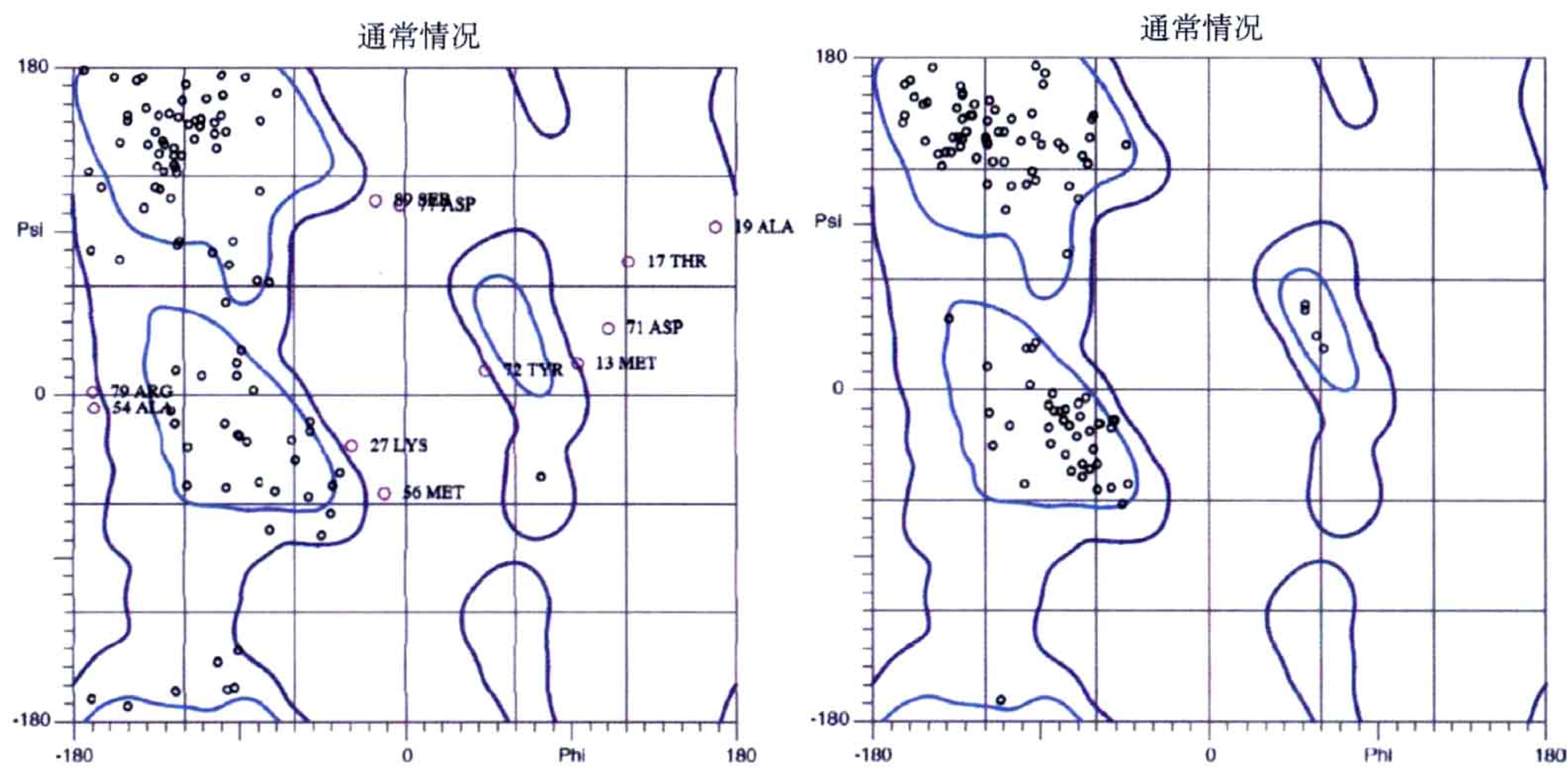


图 B.4　PBD 中同一蛋白质的两个不同模型的 Ramachandran 图谱。红色圈标注的残基超出了允许的范围。左图表明坐标是不正确的。

表 B.1　指认蛋白质和核酸的某些分子特性所需分辨率的评估

蛋白质	核酸	分辨率 /Å
大致形状	可将核酸与蛋白质区分	20
	双螺旋	12
α 螺旋	单链	9
β 折叠	堆积的碱基对磷酸基团和核糖	4
大侧链	—	3.5

续表

蛋白质	核酸	分辨率 /Å
小侧链上成形的泡状密度	可区分的嘌呤和嘧啶	3.2
侧链的构象	—	2.9
羰基，肽平面	单个碱基	2.7
	单孔的糖分子	2.4
单孔的 Pro 残基芳香环的孔	—	2.0
单个原子	单个原子	1.5

分辨率不是衡量电子密度图质量的唯一参数。被称为 R- 值的参数可以用来描述所得衍射数据对原子模型解释的优劣程度。其中最重要的是 R_{free}，它可比较当有一小部分衍射数据没有用来精修时的模型。R- 值随着分辨率的增加而改进（降低），因为模型将更加精确。

另一个能够显著提高电子密度的方法是对相同分子进行平均化。病毒分子具有大量的、具不同取向的相同分子，从而提供了对电子密度进行平均和优化的极好选择。这就使得在较低分辨率下进行蛋白质模建成为可能。

不同的 wwPDB 节点（如 RCSB、MSD、PDBj）分别提供了它们各自的搜索工具，也提供了为进一步分析而下载的工具。另外，各种设置（登录端口）也提供了到其他数据库的链接，如序列数据库、结构图示等。

2008 年 6 月，PDB 文库具有 51 000 个条目，但是其中许多条目是同一蛋白质的突变体或蛋白质与不同底物的复合物。例如，大约有 400 种不同的噬菌体 T4 溶菌酶突变体（一个通常用来研究蛋白质折叠与稳定性的例子）和 300 种 HIV 蛋白酶与不同抑制剂的复合物的相关条目。

每一条目都包含相关发表文献及测定方法的描述，以便对所得信息的质量加以评判。

B.2.2 分级的折叠数据库

大量的数据库是在收集蛋白质数据文库信息的基础上得到的。这些库包括分子图像、序列和其他的一些信息。两个数据库以分级的方式整理了已存在的结构信息。这些数据库对分析进化关系及进行功能比较非常有用。

B.2.2.1 SCOP：蛋白质的结构分类

SCOP（structural classification of protein）数据库是 A.Murzin 和他的同事一起创建的。这个数据库主要是对蛋白质的折叠进行手工分类，因此这样的分类是有主观因素的。

另外，所积累的蛋白质构象信息和特殊蛋白质的进化细节尽管对分类很有帮助，但是却不易整合到计算程序中去进行自动折叠分类。

分类的等级（图 B.5）可以分别为种类、折叠、超家族、家族和域（单个蛋白或蛋白质的结构域）。相同家族中的结构域被认为是同源的，也就是说，是从共同的祖先分化而来的。同源蛋白结构域具有相似结构的假设是基于其序列和功能的相似性。在同一超家族中的蛋白质具有相同的折叠和相关的功能，因此也可能具有共同的祖先，但是它们在序列和功能上都有太多不同，很难断定是否同源。在“折叠”这一等级中，蛋白质具有相同的拓扑结构，但是除了有限的结构相似性外，没有证据表明它们之间存在进化关系。

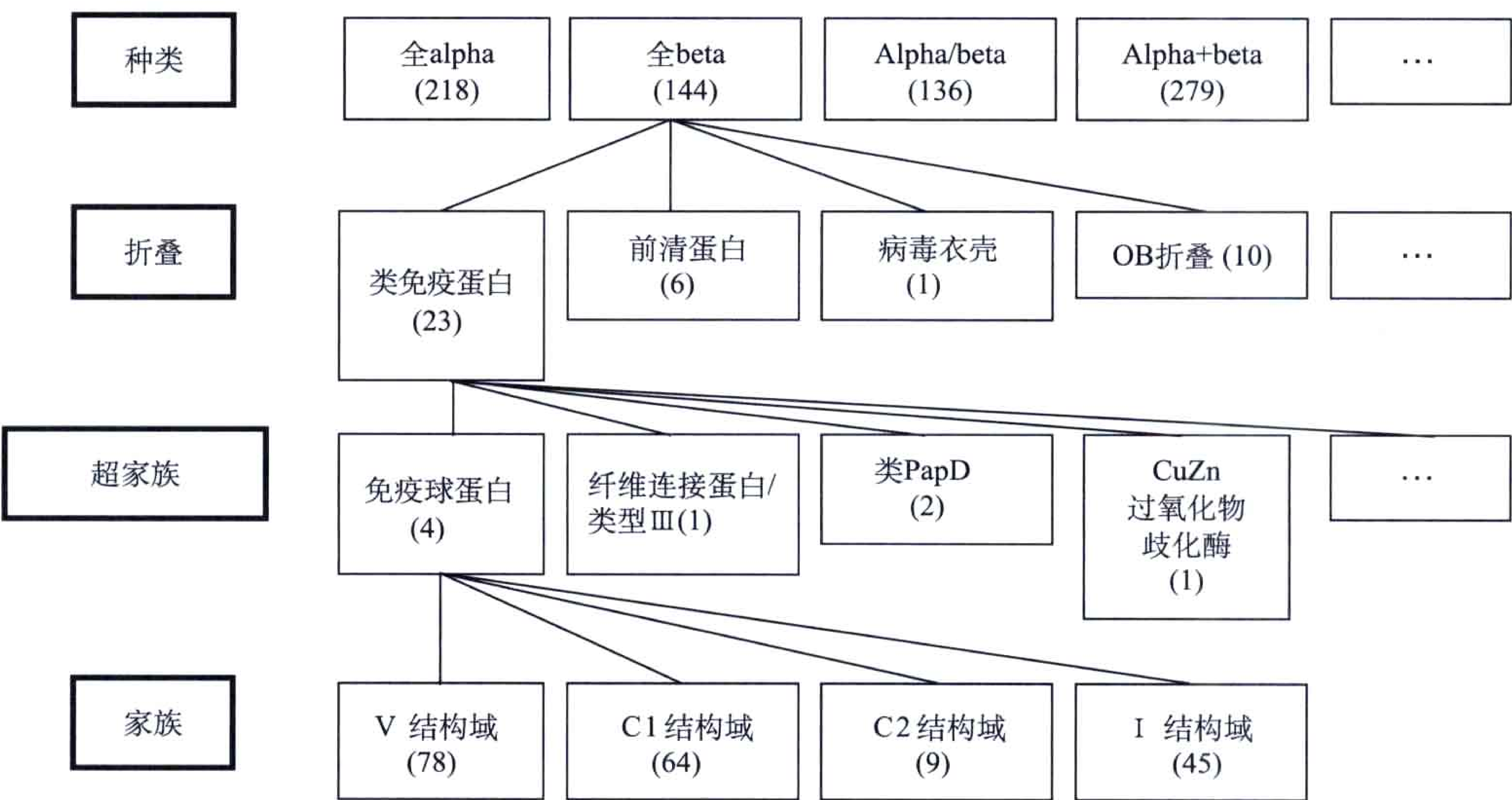


图 B.5　SCOP 分类。图例是免疫球蛋白的结构域逐级分类。括号中标明了每组中子集的数目。

SCOP 数据库是针对结构域而不是整个蛋白质分类。这就意味着多结构域蛋白被分为它的组成结构域。许多时候当一种结构域被不同的蛋白质共享时，这种分类方式是非常有用的。高质量的分类已经使得这一数据库成为研究进化关系的标准。这一数据库可以被检索并且已经连接到 PDB 的每一独立条目中。

B.2.2.2　CATH

另一个类似的数据库是 CATH，它是由 C.Orengo 和 J.Thornton 创建的。这个数据库按种类、构造、拓扑（折叠家族）、超家族及序列家族［CATH 是种类（class）、构造（architecture）、拓扑（topology）、同源超家族（homologous superfamily）的缩写，图 B.6］来分类的。

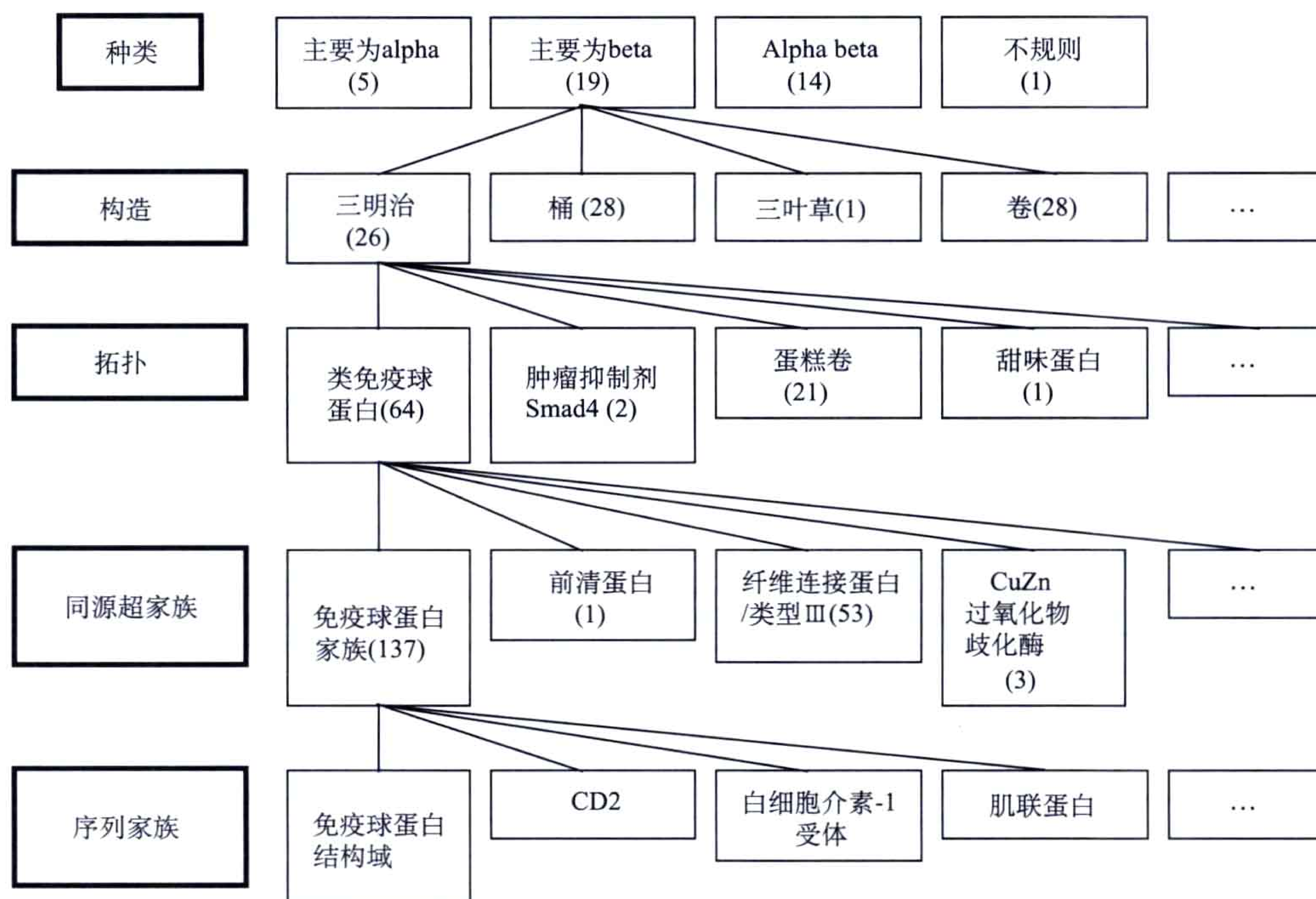


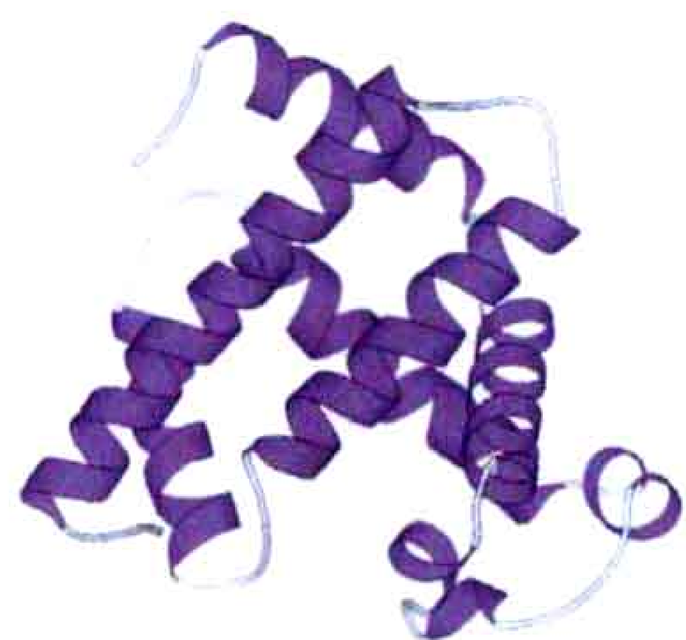
图 B.6　CATH 分类。同样以免疫球蛋白为例。值得注意的是，前清蛋白（甲状腺素）在 CATH 中被归为类免疫球蛋白拓扑类，但在 SCOP 中是一个独立的折叠（图 B.5）。

CATH 中的结构分类在很大程度上是由自动程序来完成的。这种自动程序起始于对蛋白质结构域的定义。大量能够自动完成这一任务的程序已被开发出来。CATH 中结构域的定义是从三种算法中得到的。在许多情况下如果无法由上述方法得到，则需用手动的方法定义结构域。

CATH 数据库中的拓扑在本质上与 SCOP 分类中的折叠相对应，而其构造则在 SCOP 分类中没有对应的层次。蛋白质具有相同二级结构元素，如一个 β 三明治和一个 α 螺旋，虽然可能有不同连接方式，但是仍然被定义为具有相同的构造。

由于分类步骤的不同，有些蛋白质在 SCOP 和 CATH 中的分类是不同的。这些不同之处可能源于用 CATH 方法识别拓扑时自动程序存在的差异。在两个不同的 PDB 条目中，同一蛋白质可能被指认为两个不同的拓扑。对于小蛋白来说还存在一些其他问题，对于它们的分类选择都是不明确的。

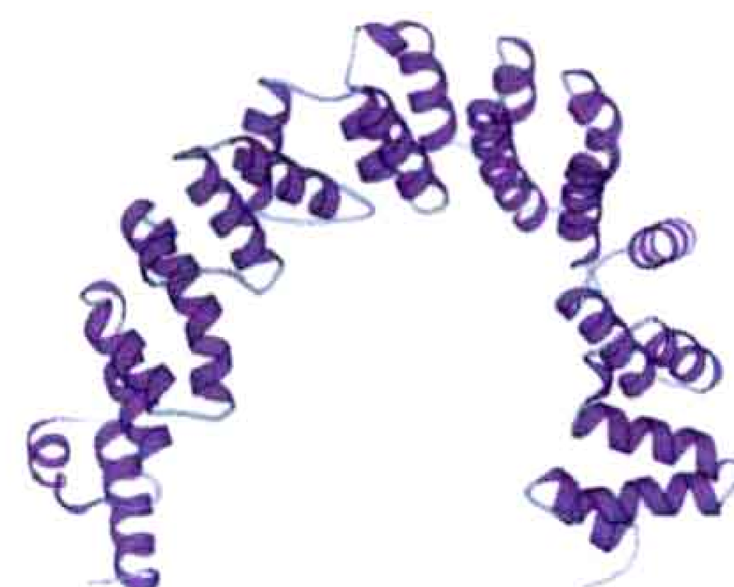
Architectures in CATH version 3.1 — α class



α 正交束（1mbn000）



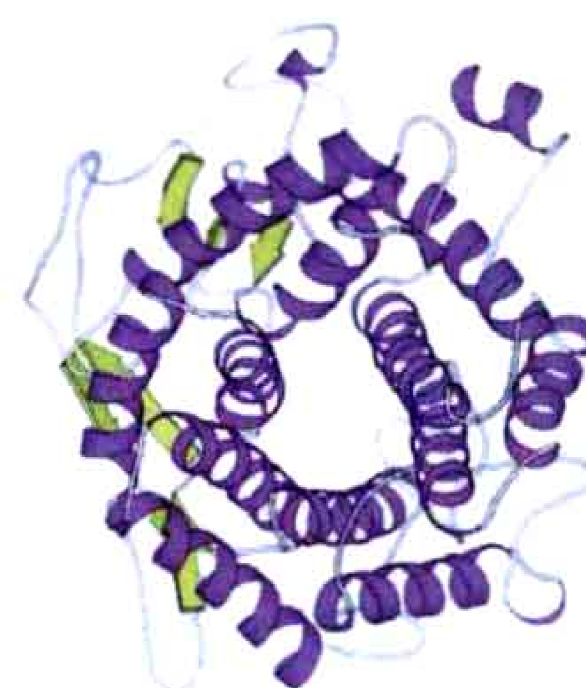
α 上下束（1e85A00）



α 马蹄（1qsaA01）

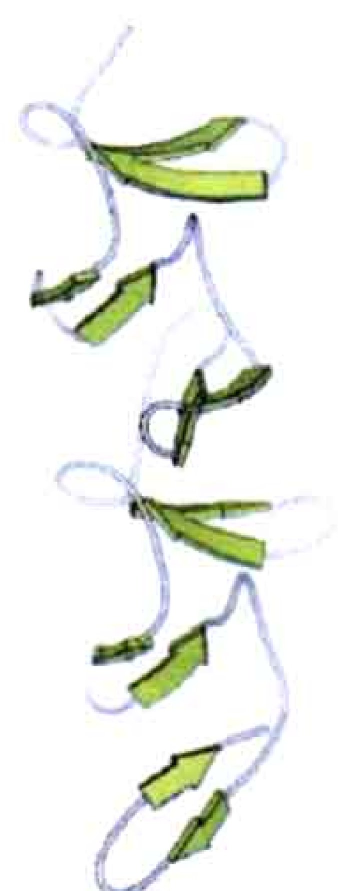


α 螺线管（1pprM01）

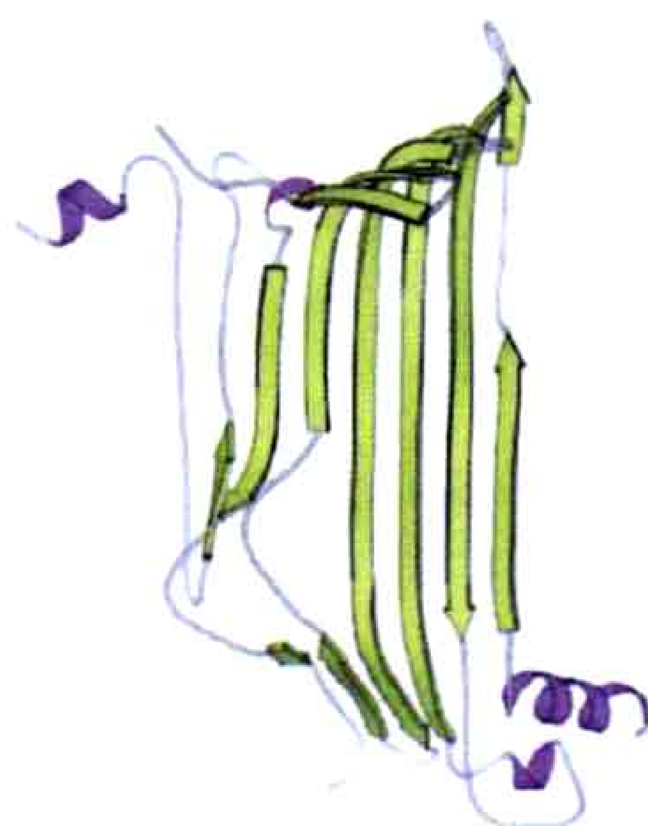


αα 桶（1fce001）

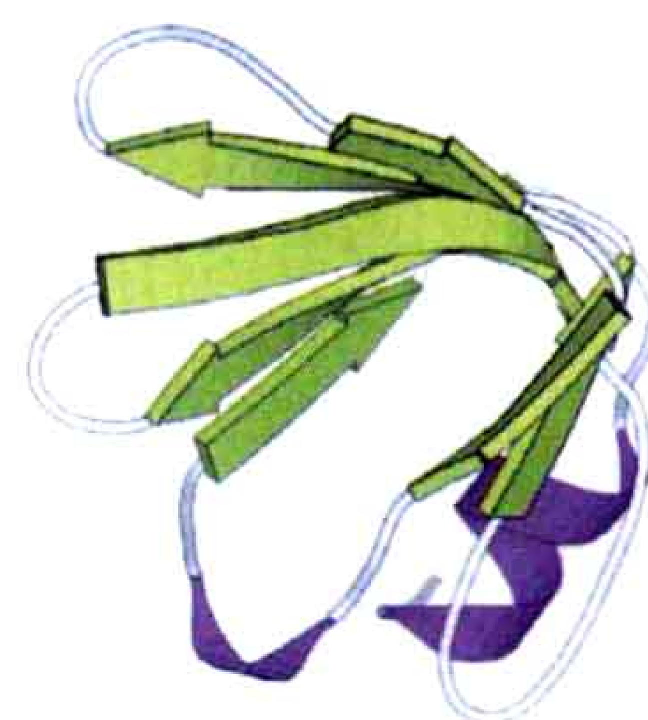
Architectures in CATH version 3.1 — β class



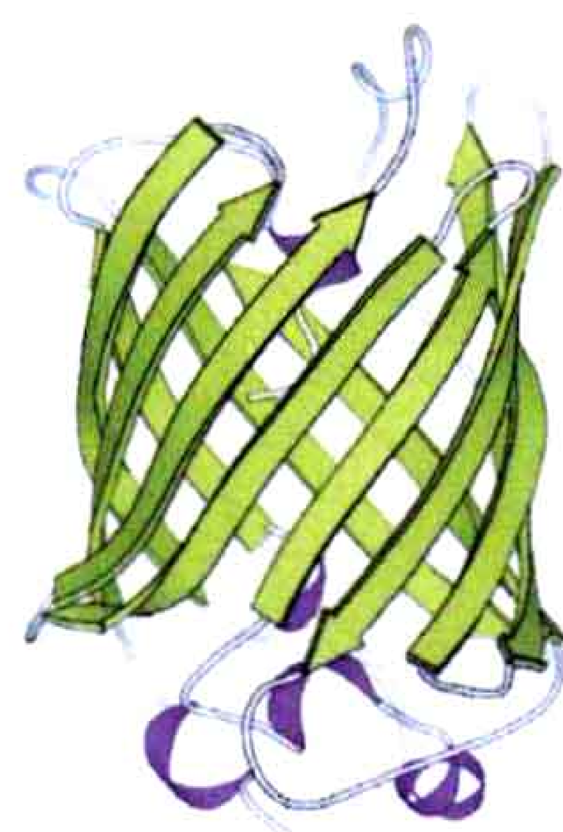
β 带子（2bmlA00）



β 片层（1lshA03）



β 卷（1h64A00）



β 桶（2fgqX00）



β 钳（4bcI000）



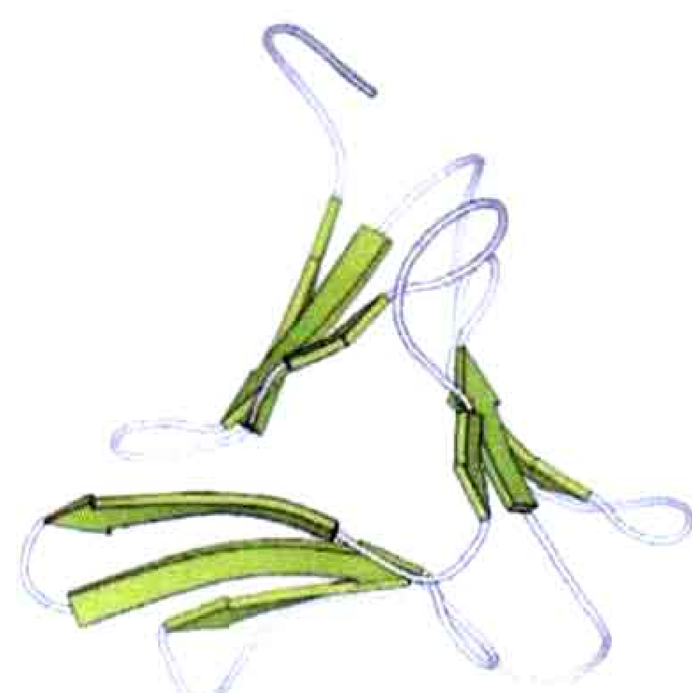
双层 β 三明治（1k5nA02）

图 B.7　CATH 分类法定义的一些构造（向 christine orengo 表示感谢）。

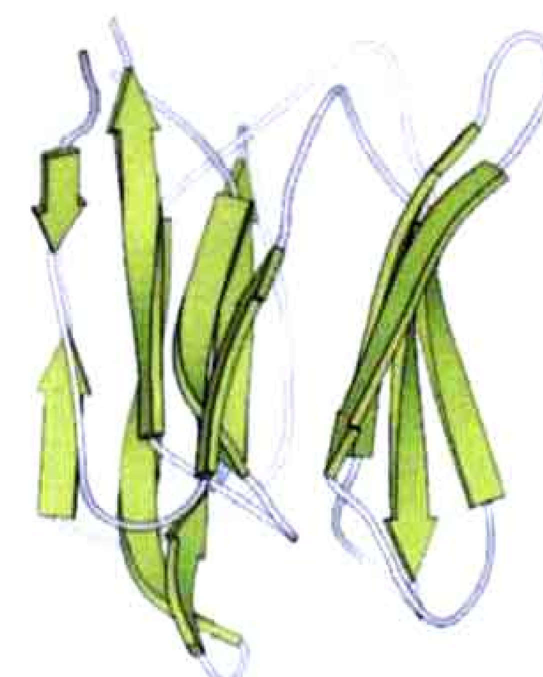
Architectures in CATH version 3.1 — β class



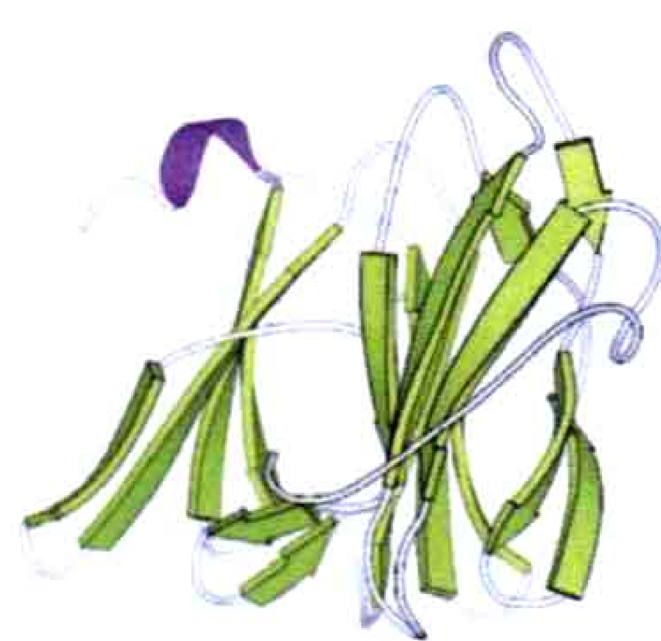
β 三叶草（1ybiA01）



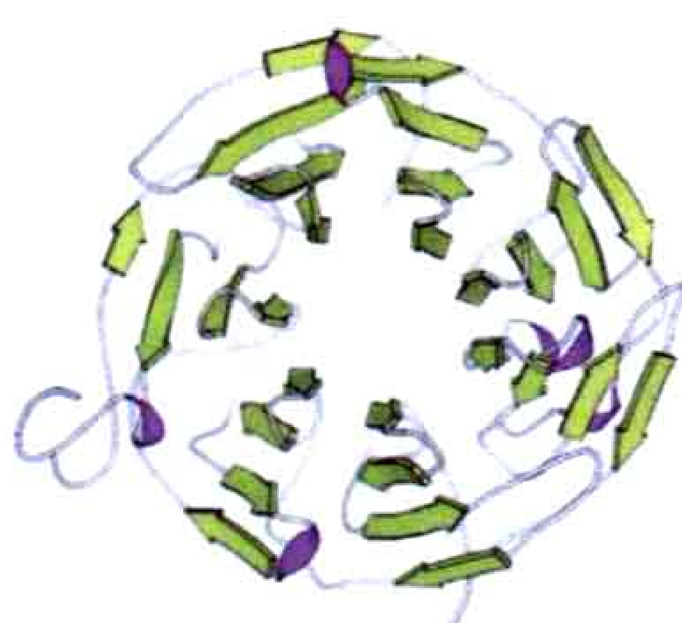
β 正交 β 棱柱（1b2pA00）



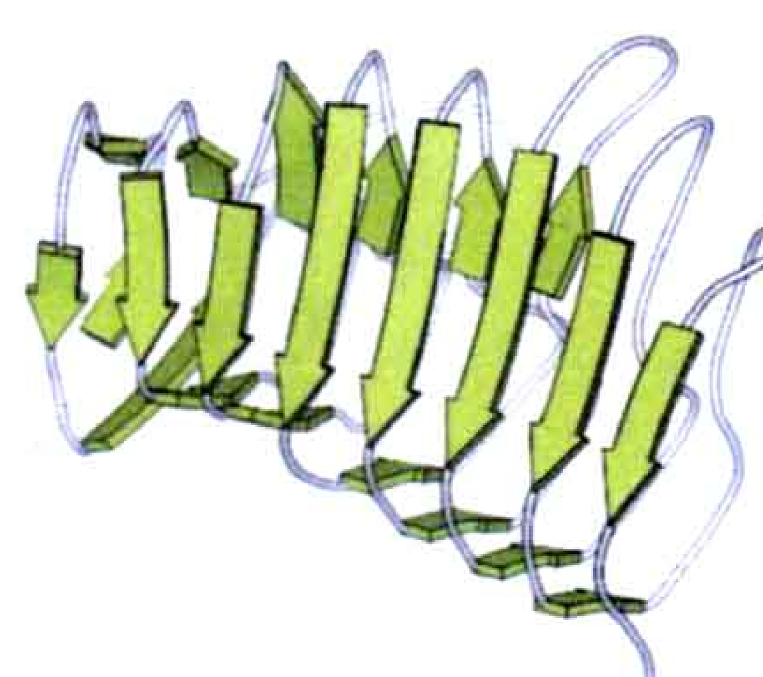
β 平行 β 棱柱（1ouwA00）



三层 β 三明治（1tg7A02）



β 螺旋桨（1k3iA02）

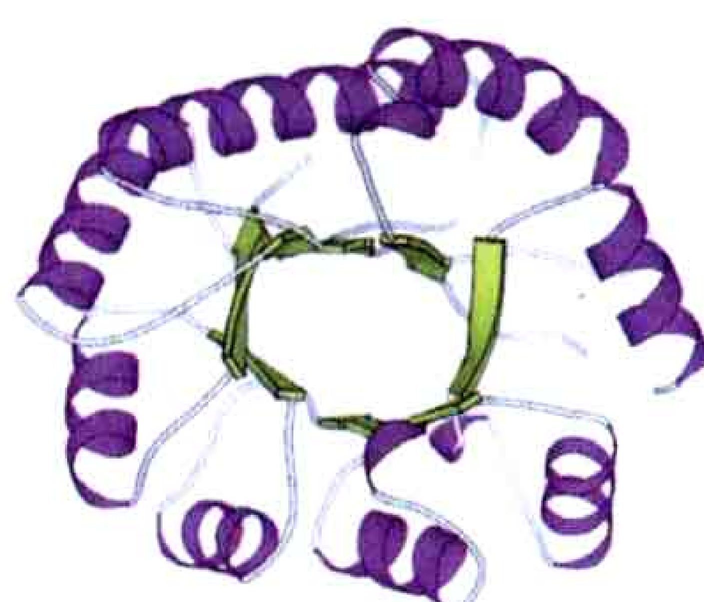


β 螺线管（1ee6A00）

Architectures in CATH version 3.1 — αβ class



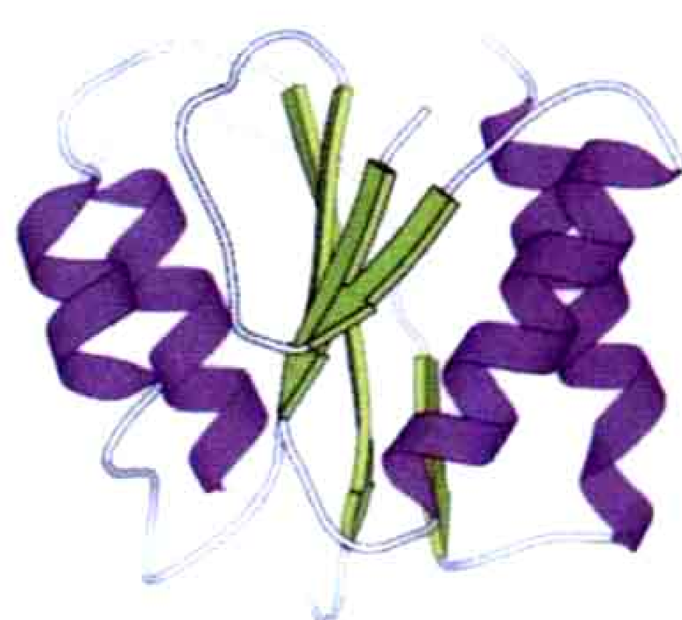
αβ 卷（1aarA00）



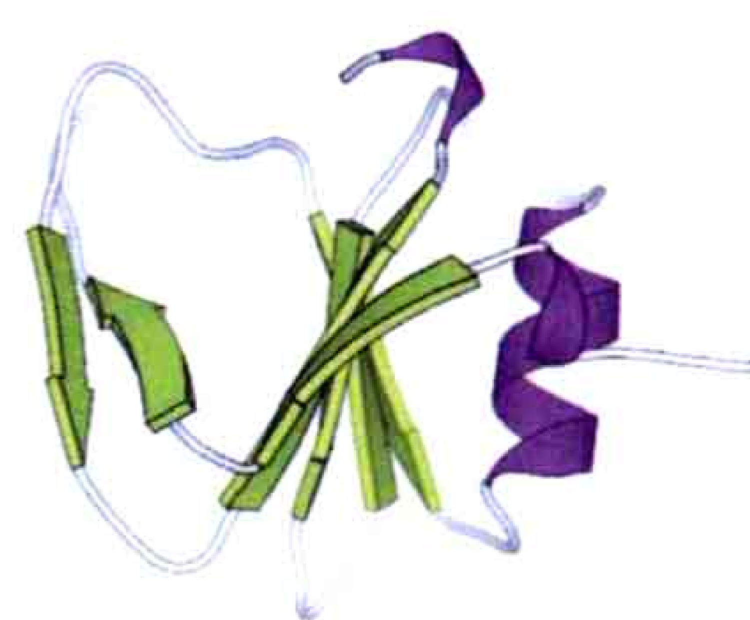
αβ 桶（7odcA02）



双层（αβ）三明治（1ay7B00）



三层（αβα）三明治（4fxn000）



三层（ββα）三明治（1bhtB01）

图 B.7 （续）

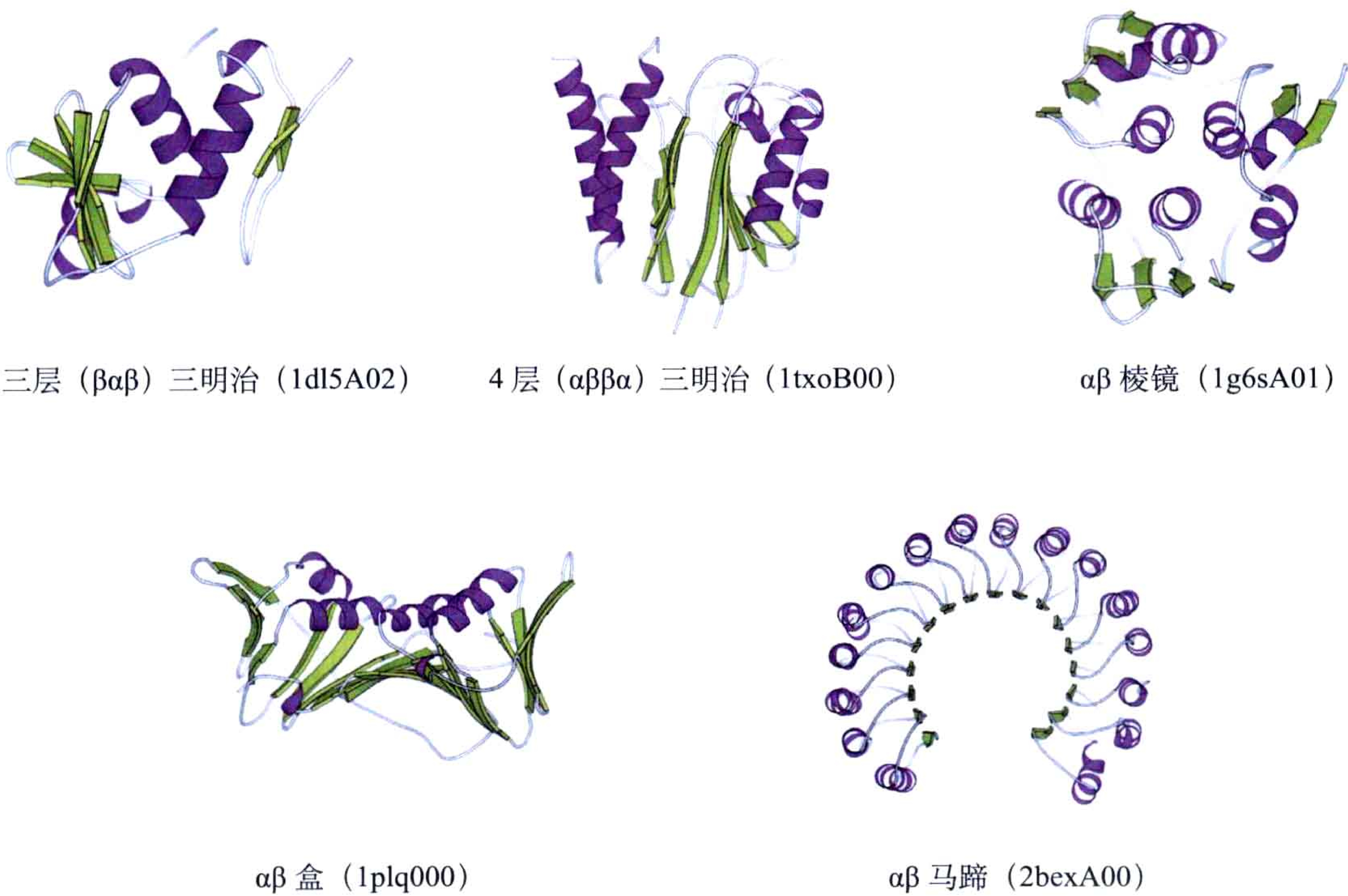


图 B.7　（续）

B.2.3　其他结构分类

也有一些结构数据库不通过分级归类，而是基于已知结构自动归类。Dali 就是这样一个数据库，它是通过 Dali 程序来比对并将蛋白质折叠分类的（表 B.2）。

表 B.2　与己糖激酶结构相近的蛋白质

Chain[a]	Z-score[b]	%ID[c]	LALI[d]	RMSD[e]	Description
1qhaA	69.1	100	903	0.0	己糖激酶（人）
1bg3B	62.4	92	883	0.9	己糖激酶（鼠）
1bdg	55.4	45	439	1.5	己糖激酶（血吸虫）
1v4sA	55.2	54	444	1.5	葡萄糖激酶异形体 2
1hkg	40.7	15	432	2.3	己糖激酶（酵母）
1ig8A	39.2	34	438	3.4	己糖激酶 PⅡ（酵母）
1xc3A	19.8	10	265	3.1	假果糖激酶

续表

Chain[a]	Z-score[b]	%ID[c]	LALI[d]	RMSD[e]	Description
1q18A	18.5	13	280	4.5	葡萄糖激酶
1woqA	17.7	16	239	3.0	无机多磷酸/ATP 葡萄糖甘露糖激酶
1e4gT	14.8	10	261	4.0	细胞分化蛋白 FtsA（热海栖热袍菌）
1yagA	14.8	9	266	4.1	肌动蛋白（酵母）
2btfA	14.7	9	269	4.1	肌动蛋白（牛）
1jcfA	14.7	11	258	3.6	棒状确定蛋白 MreB（热海栖热袍菌）
1dkgD	14.6	10	267	3.9	核苷酸交换因子（大肠杆菌）
1s3xA	14.1	10	269	4.3	HSP70（人）
1mwkA	13.5	10	250	3.5	质粒隔离蛋白（大肠杆菌）
1huxA	13.3	14	224	3.8	羟基戊二酰辅酶 A 脱氢酶激活剂（发酵支原体）
1nbwA	12.2	10	265	4.0	甘油脱氢酶激活酶 α 亚基
1tuuA	11.7	10	252	4.2	乙酸激酶
1g1fY	10.6	14	245	3.7	甘油激酶
1hjrA	6.2	5	123	4.0	Holliday 结点、解离酶
1c0mC	4.3	5	87	2.9	整合酶（Rous 肿瘤病毒）

a PDB 代码附加链的名称（A,B,…）。表中列出了与 1QHA（己糖激酶）链 A 结构相近的蛋白质。

b Z- 分值表明结构的相似性。这一列是简略的。位于列顶部的是几个己糖激酶的结构。相对较高的分值也出现在肌动蛋白（见 12.1 节）、与肌动蛋白相关的蛋白质（MreB、FtsA、ParM）及热激蛋白（HSP70, 见 9.1.2.3 节）。以上两组蛋白质都与己糖激酶在序列上出入很大但在结构上却有几分相似。它们之间共同的功能（ATP 结合与水解）说明了其在进化上的相关性。位于列中的最后两个链是作为与蛋白质仅有局部结构相似性的例子来展示的。

c 残基比对中相似性的百分比。

d 比对残基的数目。

e 叠加后与第 q 条链上 Cα 原子的距离的均方根（Å）。

与 CATH 和 SCOP 的程序不同，Dali 的程序是完全自动的。对于一个序列相似性小于 90% 的子集的所有结构来说，Dali 数据库列出了 PDB 中所有的结构相似的分子。其中所含信息包括某一给定结构与其他条目间的结构和序列的相似程度。通过由 Z- 分值来评判的结构相似性，所列蛋白质可进一步被分类。Z- 分值是标准偏差和尺寸相似的随机蛋白平均结构相似性的比值。Z- 分值小于 2.0 的蛋白质不被认为是相似的，从而不被列出。但是也不存在一个明确的 Z- 分值作为评判进化关系的界限。由于 Dali 所列蛋白质纯粹是基于计算所得的分值相似性，因此其功能相似性便可作为一重要的附加因素来评判进化关系。

NCBI 通过 VAST（vector alignment search tool）程序来比较结构。它的结构数据库 MMDB 包含与所有 PDB 条目相关的结构。

VAST 和 Dali 及其他折叠比对的程序可作为服务器来将未知结构与数据库中的结

构相比较。

延伸阅读

Holm L, Sander C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**: 123-138.

Murzin AG, Brenner SE, Hubbard T, Chothia C. (1995) SCOP: Astructural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**:536-540.

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. (1997) CATH —A hierarchic classification of protein domain structures. *Structure* **5**: 1093-1108.

数据库和服务器的链接

PDB at RCSB: http://www.rcsb.org/pdb/home/home.do

PDB at MSD: http://www.ebi.ac.uk/msd/

PDBj: http://www.pdbj.org/

PDBsum: http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/

SCOP: http://scop.mrc-lmb.cam.ac.uk/scop/

CATH: http://www.biochem.ucl.ac.uk/bsm/cath/

Dali: http://ekhidna.biocenter.helsinki.fi/dali/start

NCBI structural database MMDB: http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml

VAST: http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml

附录 C

蛋白质构象的预测

C.1 二级结构的预测

C.1.1 二级结构预测方法的基本原理

蛋白质的序列包含着所有指导蛋白质折叠的必要信息，这一认知激励着科学家们试图去从序列中直接预测蛋白质的构象。经过数十年的努力，尽管方法在不断地改进，但是这一目标仍然没有实现，原因主要有两点。第一，即使对一个很小的蛋白质来说，也存在着大量可能的构象。即便使用现代化的计算机并且将任务简化到只对每个氨基酸残基的一小部分排列予以考虑，也很难将所有的构象都展现出来。第二，体系的自由能（蛋白质和周围溶剂间）将不得不通过简化的势能方程计算得到，这样就不能准确地将体系中所有可能的作用都考虑在内，从而得到正确的最小值。此外，折叠的构象只是在外围比非折叠的状态更加稳定。

早期，人们认识到这个问题可以通过将预测分为两个阶段来进行简化：首先将蛋白质的二级结构预测出来，然后将折叠束或是螺旋装配成正确的折叠。从某种程度上讲，这一方法模拟了一种可能的蛋白质折叠机理，也就是首先形成的是局部的二级结构，然后将它们排布为折叠构象。

蛋白质多肽片段二级结构的折叠是由氨基酸序列决定的。利用氨基酸具有偏好某一种构象的趋势来进行预测便是一种办法。基于氨基酸残基形成 β 束、螺旋或是转角的倾向性来预测二级结构的方法已经被发展起来。这种倾向性已经通过研究溶液中的小肽构象，或是通过对特定残基在已知结构蛋白质不同类型的二级结构中的统计分析被总结出来。

C.1.2 预测方法

C.1.2.1 Chou-Fasman 法

第一个被广泛用来预测二级结构的方法是 Chou-Fasman 方法。这一方法基于所有

20 种氨基酸形成 α 螺旋和 β 束的倾向性，其中包括螺旋的形成或断开以及 β 束的形成或断开（表 C.1）。这些倾向性是在对现有已知结构的蛋白二级结构中不同残基的存在形式进行统计学分析的基础上得到的。这一方法包含了许多形成和断开螺旋及 β 束的规则。对于 6 个残基（螺旋）或是 5 个残基（β 束）界定的范围来说，如果其倾向性大于某一特定的阈值，这一片段就会形成其对应的二级结构。这个螺旋或是 β 束会不断地延伸，直到遇到 4 个残基所界定范围的倾向性小于某一阈值为止。当结合一些其他的规则时，如当螺旋和 β 束同时被预测出来时该如何做出取舍，就给出了一套从螺旋、β 束或者环状结构预测二级结构的办法，这里的环状结构代表所有不是螺旋或者 β 束的构象。简单的 Chou-Fasman 预测方法可以通过纸笔来完成，不过通常是通过程序来进行的。

表 C.1　Chou-Fasman 方法中形成 α 螺旋、β 束以及转角的倾向性

螺旋			β 束			转角	
Glu	1.51	H	Val	1.70	H	Asn	1.56
Met	1.45	H	Ile	1.60	H	Gly	1.56
Ala	1.42	H	Tyr	1.47	H	Pro	1.52
Leu	1.21	H	Phe	1.38	h	Asp	1.46
Lys	1.16	h	Trp	1.37	h	Ser	1.43
Phe	1.13	h	Leu	1.30	h	Cys	1.19
Gln	1.11	h	Cys	1.19	h	Tyr	1.14
Trp	1.08	h	Thr	1.19	h	Lys	1.01
Ile	1.08	h	Gln	1.10	h	Gln	0.98
Val	1.06	h	Met	1.05	h	Thr	0.96
Asp	1.01	I	Trp	0.93	i	Trp	0.96
His	1.00	I	Asn	0.89	i	Arg	0.95
Arg	0.98	i	His	0.87	i	His	0.95
Thr	0.83	i	Ala	0.83	i	Glu	0.74
Ser	0.77	i	Ser	0.75	b	Ala	0.66
Cys	0.70	i	Gly	0.75	b	Met	0.60
Tyr	0.69	b	Lys	0.74	b	Phe	0.60
Asn	0.67	b	Pro	0.55	b	leu	0.59
Pro	0.57	B	Asp	0.54	b	Val	0.50
Gly	0.57	B	Glu	0.37	b	Asn	1.56

C.1.2.2　GOR 法

一个较为复杂的方法是 Garnier-Osguthorpe-Robson(GOR) 法，它包含稍长的多肽链片段的信息。对于螺旋、片层和转角来讲，其构型的取向不是由单个的残基决定的，而是由所有的残基类型决定的。对于序列中的每一个位点 j 来说，其两侧的 8 个残基都会被考虑在内。对于一个螺旋特征表（图 C.1）来说，其信息包含了每一种氨基酸类型

出现在序列中这 17 个位点的可能性，以用于判断 j 位点是否为螺旋。这样，螺旋特征表就有 20 × 17 个条目。在位点 j 所预测的构象状态（α 螺旋、片层或转角）就通过计算位点 j 周围 17 个残基范围内所有残基与位置相关的倾向性的总和。其倾向性表反映出脯氨酸往往出现在螺旋的 N 端，而不是 C 端（–1 位的 Pro 的贡献是 –140，而 +1 的 Pro 的贡献是 –60）。同时也反映出带正电的残基通常出现在螺旋的 C 端，而带负电的残基则出现在 N 端的趋势（2.1 节）。GOR 法已经被程序化并且作为标准方法使用许多年。

	C -8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	N 8
gly	-5	-10	-15	-20	-30	-40	-50	-60	-86	-60	-50	-40	-30	-20	-15	-10	-5
ala	5	10	15	20	30	40	50	60	65	60	50	40	30	20	15	10	5
val	0	0	0	0	0	0	5	10	14	10	5	0	0	0	0	0	0
leu	0	5	10	15	20	25	28	30	32	30	28	25	20	15	10	5	0
ile	5	10	15	20	25	20	15	10	6	0	-10	-15	-20	-25	-20	-10	-5
ser	0	-5	-10	-15	-20	-25	-30	-35	-39	-35	-30	-25	-20	-15	-10	-5	0
thr	0	0	0	-5	-10	-15	-20	-25	-26	-25	-20	-15	-10	-5	0	0	0
asp	0	-5	-10	-15	-20	-15	-10	0	5	10	15	20	20	20	15	10	5
glu	0	0	0	0	10	20	60	70	78	78	78	78	78	70	60	40	20
asn	0	0	0	0	-10	-20	-30	-40	-51	-40	-30	-20	-10	0	0	0	0
gln	0	0	0	0	5	10	20	20	10	-10	-20	-20	-10	-5	0	0	0
lys	20	40	50	55	60	60	50	30	23	10	5	0	0	0	0	0	0
his	10	20	30	40	50	50	50	30	12	-20	-10	0	0	0	0	0	0
arg	0	0	0	0	0	0	0	0	-9	-15	-20	-30	-40	-50	-50	-30	-10
phe	0	0	0	0	0	5	10	15	16	15	10	5	0	0	0	0	0
tyr	-5	-10	-15	-20	-25	-30	-35	-40	-45	-40	-35	-30	-25	-20	-15	-10	-5
trp	-10	-20	-40	-50	-50	-10	0	10	12	10	0	-10	-50	-50	-40	-20	-10
cys	0	0	0	0	0	0	-5	-10	-13	-10	-5	0	0	0	0	0	0
met	10	20	25	30	35	40	45	50	53	50	45	40	35	30	25	20	10
pro	-10	-20	-40	-60	-80	-100	-120	-140	-77	-60	-30	-20	-10	0	0	0	0

图 C.1　GOR 方法的螺旋特征表。位于中间的“0”列表示当前位点以及处于该位点的氨基酸的特性。那些处于该残基前后的残基贡献在相应的列中有所体现，并且被相加给出当前位点的总螺旋分值。例如，标有 − 6 的那一列包含了所有在当前残基 C 端第 6 个位点处不同残基的贡献。摘自 Garnier J, Osguthorpe DJ, Robson B. (1978) *J Mol Biol* **120:** 97-120。

C.1.2.3　神经网络法

有许多的方法具有比上述提到的两种方法更准确的预测能力。这种预测能力的提高在一定程度上是基于应用多序列比对来进行预测。在同源蛋白中取代和插入 / 删除的方式包含着关于蛋白质折叠特性的重要的信息。例如，在二级结构的元件中，插入和删除是不常见的，而保守的甘氨酸和脯氨酸倾向于在转角中出现，等等。

一个这样程序的例子是 PHDsec，它是同过一个双层的神经网络来预测二级结构的。

一个神经网络就是一个能够建立不同类型输入和输出信息关系的计算机认知程序。这些信息在程序中以参数（权重）的形式储存（图 C.2）。当用神经网落来预测二级结构时，其目的就是要将序列的信息与观察到的它们相对应的二级结构联系起来。这个网落包含了一系列的层（输入层、隐藏层、输出层），信息将以输送的方式从其中一层流向另一层。可以用已知结构的同源序列比对来“训练”这一网络。在这一过程中，序列和所观察到的二级结构的关系被用来进行参数计算。这些参数暗示了特定的残基与某种二级结构相对应的可能性，因此可以被用来预测结构未知的蛋白比对序列的二级结构。

PHD 神经网络中（图 C.3）包含一个将序列信息与二级结构连接起来的层（序列 -

结构层）。在训练过程中，将每 13 个氨基酸的范围里，不同残基的发生事件与二级结构中处于中心位置的残基联系起来。与图 C.2 相似，其参数将反映二级结构的特点。在接下来的步骤中（结构 - 结构层），由上一层输出的界定在每 17 个残基范围内的参数被用来预测中心残基的二级结构。在这里，应该训练网络不去预测那些实际上不存在的短片段的二级结构。

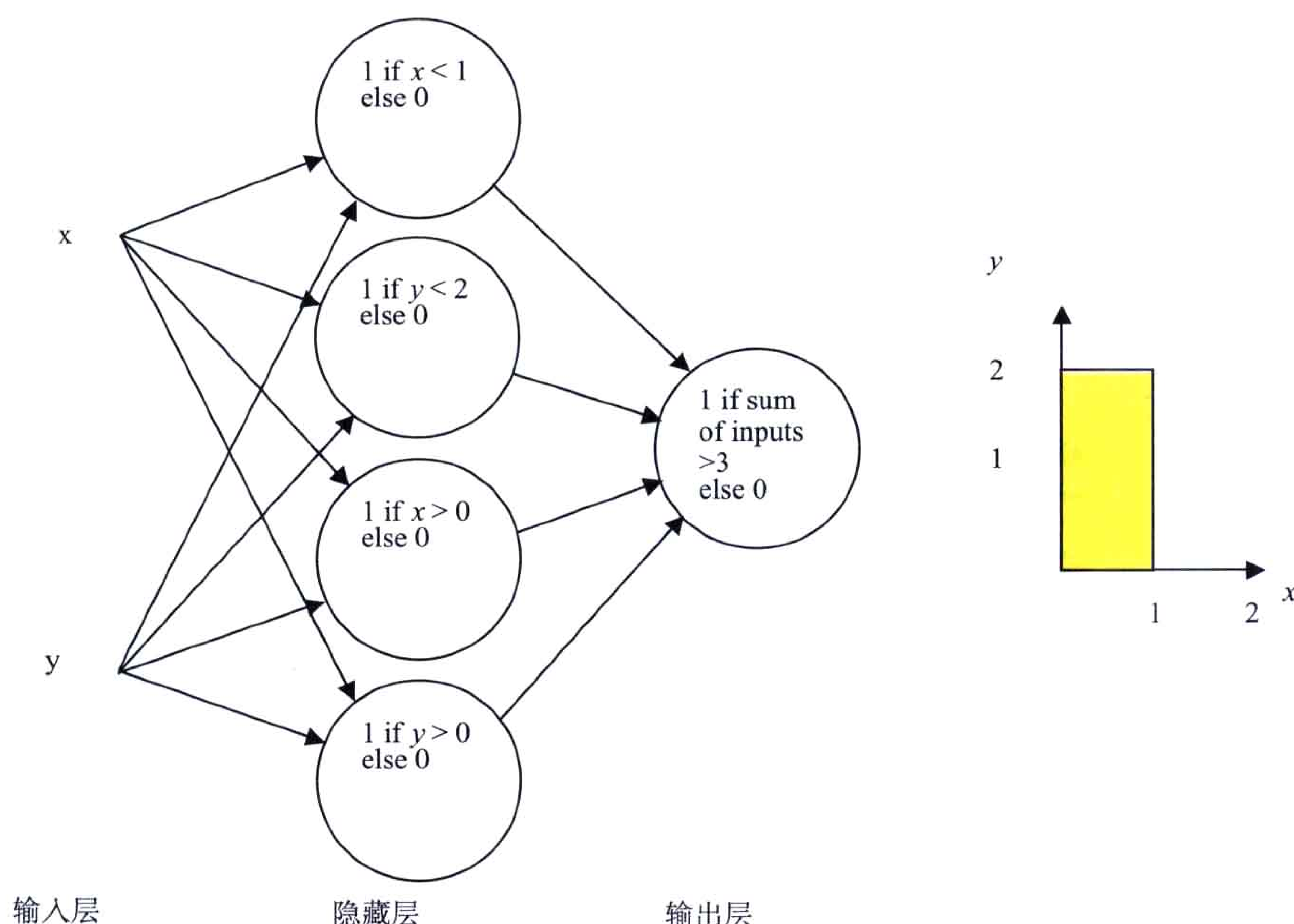


图 C.2　一个简单的神经网络图，可以用来决定一个点是否落在阴影部分。如果输出层产生的数字是 1，则该点落在阴影区内 [摘自 Lesk AM.(2002)*Introduction to Bioinformatics*. Oxford University Press, Oxford]。

当神经网络被几个已知结构的蛋白质训练完成后，这个网络就可以用来预测二级结构。PHDsec 的一个主要特性是首先用搜寻序列去寻找可与其比对的同源序列（运用 BLAST 序列比对工具）。所得比对序列被输入网络中进行预测。尽管在网络中有更加精细的处理，对上述方法的主要改进还是基于序列的比对。例如，那些在比对序列中有删除或是插入的位置往往会出现在环状区域。当用 PHDsec 方法预测单个序列时，其准确度会显著下降，仅比一些旧方法好一些。

C.1.2.4　正确性评估

评估二级结构预测方法正确性的标准方法是预测一系列已知结构蛋白的二级结构并且计算出其 Q 值，也就是蛋白残基被正确归属的百分比。对于一个完整的随机预测，如果蛋白质具有相等数量的三种类型（螺旋、片层及其他类型）的构象，则其 Q 值将会是 33%。 GOR 法的准确性大约为 60% 左右，而 Chou-Fasman 法的准确性则相对低一些。基于神经网络和其他方法，当有多条序列存在时，最好的预测程序的准确性可达到 80%。

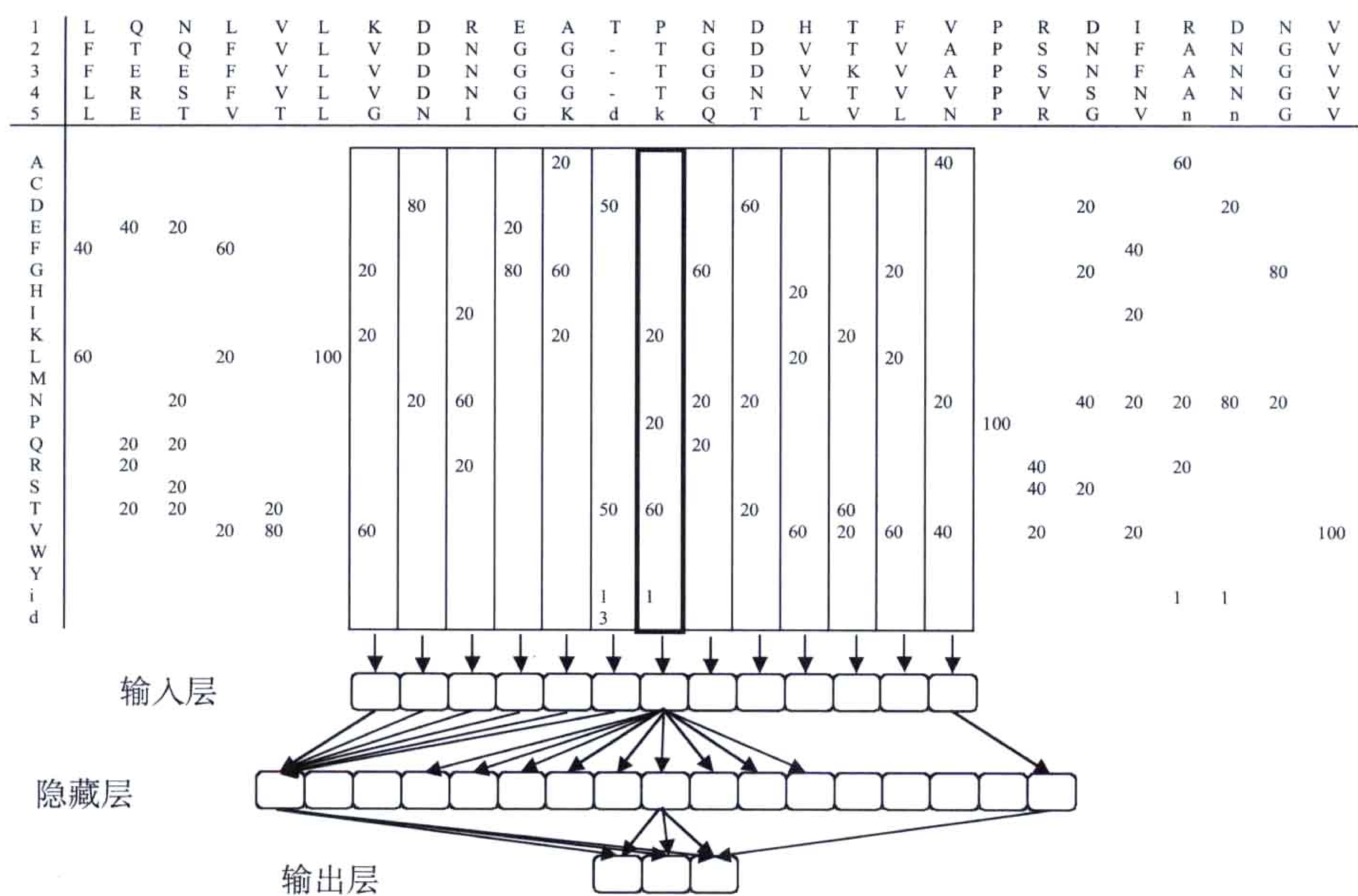


图 C.3　PHDsec 中部分神经网络的示意图。蛋白质与 4 个相关的序列相比对，在由 13 个残基组成的窗口中的序列信息被输入 13 × 24 = 312 个单元组成的输入层中（每个单元对应一个氨基酸，或是末端或间隙）。一个由 17 个节点组成的隐藏层将输入层和输出层联系起来，每一个节点对应于螺旋、折叠束和环的可能性。第二个网络被用来改进从该步骤所得的局部预测。

有些情况下，许多二级结构元件可以被准确地预测出来，但是对它们起始和结尾部分预测的准确性则不高。事实上，即使结构已知，对二级结构元件的起始与末尾的界定也是含糊不清的。基于实验测定的三维结构，一个最常用来定义蛋白质二级结构的程序是 DSSP，这一程序已被用来定义了许多不同类型的二级结构，包括三种不同类型的螺旋。这一程序主要是基于对氢键键能的估计。

目前，最好的二级结构预测方法已相对准确，不太可能进一步优化它们了。二级结构的预测方法往往是基于其局域序列，而真实的折叠和二级结构则依赖于非局域的相互作用。蛋白质二级结构中主要依赖于非局域相互作用的部分往往是程序所不能正确预测的部分。有实例表明由最多 8 个残基组成的相同序列形成完全不同的结构，这说明非局域影响对多肽构象的形成至关重要（多变序列；图 C.4）。

图 C.4　相同的八肽（GSLVALGF）可以分别在磷脂酰 3- 激酶（PDB：1PHT）（左图）和胰凝乳蛋白酶抑制剂（PDB：1WBC）（右图）中具有 α 螺旋和 β 片层的构象。

对一个特定预测的质量做出评估是非常有用的，只是还没有好的方法能达到这一目的。

C.2 其他局部特性的预测

C.2.1 跨膜信号螺旋结构的预测

膜蛋白在许多细胞功能中起着重要的作用。由于它们在磷脂双层膜中的定位，因而具有特殊的性质。在大多数真核细胞膜蛋白中，α 螺旋横贯磷脂双层的疏水区域。对一些蛋白质，单个螺旋可将原本暴露在胞内或胞外的部分固定在膜上。对另一些蛋白质，则需多个跨膜螺旋，连接它们的是短环或者能够形成独立的膜外结构域的较长片段。细菌的孔蛋白是个特例，它的跨膜区是由反平行的 β 束构成的柱状体。跨膜（TM）螺旋处于一个偏好非极性侧链的环境中，因此，这样的螺旋就可以简单地通过寻找连续的非极性残基来得到。对于与膜表面垂直的螺旋来说，大约需要 20 个残基才能完成跨越细胞膜的疏水部分，如牛视网膜色素（图 C.5 和图 C.6）。一个传统的方法是在 20 个残基的范围内计算它们的平均疏水性，这一方法已被成功用来预测许多的跨膜螺旋。对于上述计算来说，基于氨基酸物理化学特性的疏水程度可以被使用。

```
  1 MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML GFPINFLTLY VTVQHKKLRT
       EE SS E ESS  TTS     TTTS  TT TTTHHHHHHH HHHHHHHHHH HHHHHHHHHH HHHSTT   S

 71 PLNYILLNLA VADLFMVFGG FTTTLYTSLH GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC
    HHHHHHTHHH HHHHHHHHHT HHHHHHHHHH TS TTHHHHH HHHHHTTHHH HHHHHHHHHH HHHHHHHHHS

141 KPMSNFRFGE NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN ESFVIYMFVV
      SSS    H HHHHHHHHHH HHHHHHHHTH HHHTSS EEE ETTTTEEEE   S   TTTTH HHHTHHHHHH

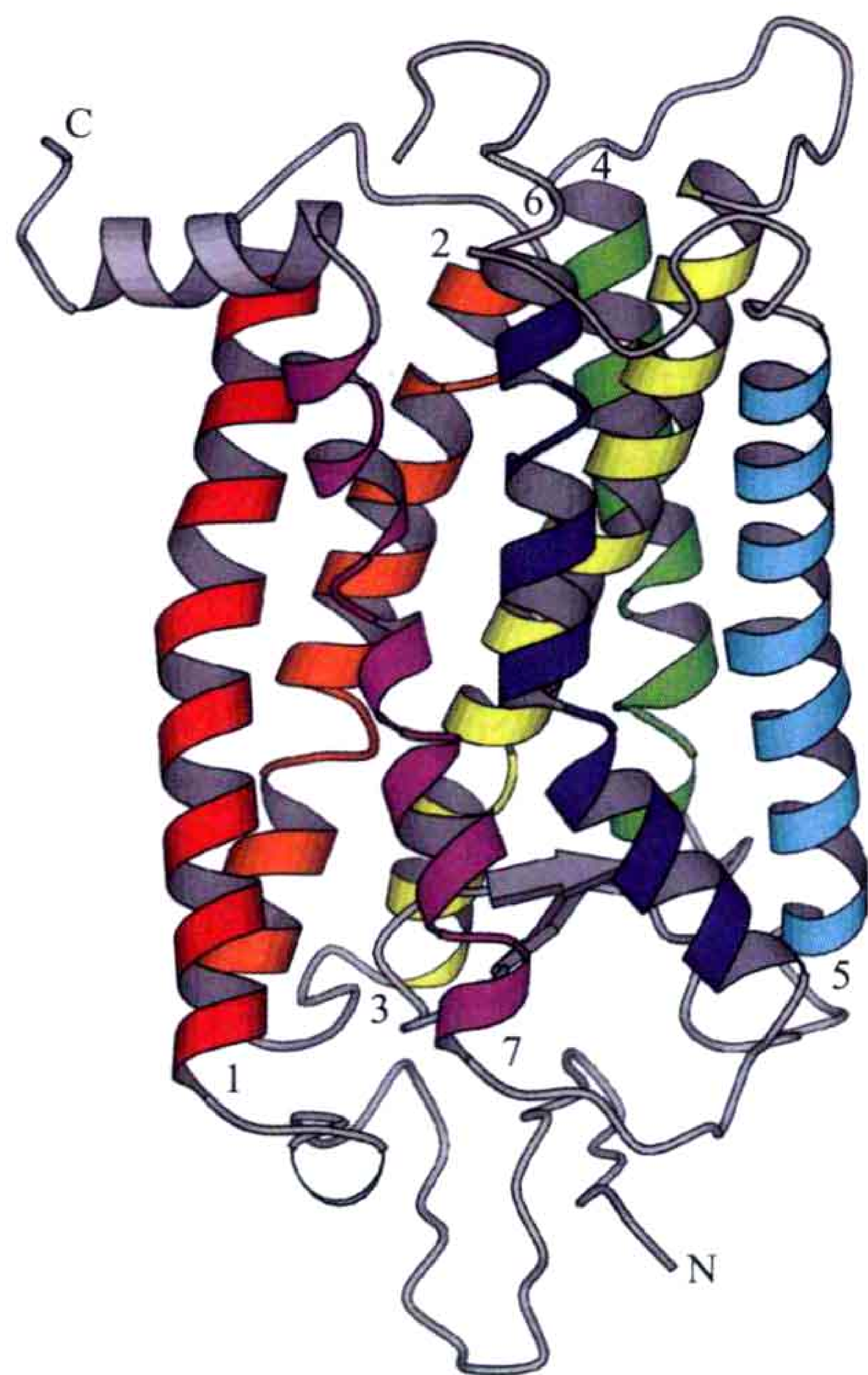
211 HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV TRMVIIMVIA FLICWLPYAG VAFYIFTHQG
    TTHHHHHHHH HHHHHTTSS   S          SSSTTHHHH HHHHTHHHHH HHHHHTHHHH HHHHHHHSTT

281 SDFGPIFMTI PAFFAKTSAV YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
        HHHHHH HHHHHGGGGT HHHHHHHHHT HHHHHHHHHH TTTT              SSS SS
```

图 C.5 牛视网膜色素的氨基酸序列和实验测定的二级结构（见图 C.6）。H 代表螺旋，E 代表延伸的 β 折叠，T 和 S 代表转角。二级结构可以与图 C.7 中预测的跨膜螺旋相对比。

许多可以准确预测跨膜片段的方法已经得到发展。这些方法可以计算出不同残基的位于跨膜及跨膜前后区域的倾向性。例如，带正电的赖氨酸和精氨酸残基通常出现在胞内而不是膜的外表面的环状结构中（“内层正电”原则）。这一原则可以被用来发现跨膜螺旋从膜外到膜内的方向（图 C.7）。

当若干个同源序列已知时，多序列的比对可以被用来提高预测准确度。这些方法可以被用来识别 90% ~ 95% 的跨膜片段，仅有少量的片段不能被正确预测。



图C.6 牛视网膜色素的三维结构(PDB: 1F88)。分子中间的部分是跨膜的部分。7个螺旋被标记出来(1: 34 ~ 64,2: 71 ~ 100,3: 106 ~ 139，4: 150 ~ 173，5： 199 ~ 225,6： 249 ~ 277，7： 285 ~ 309) 。所有的螺旋有轻微的扭曲。

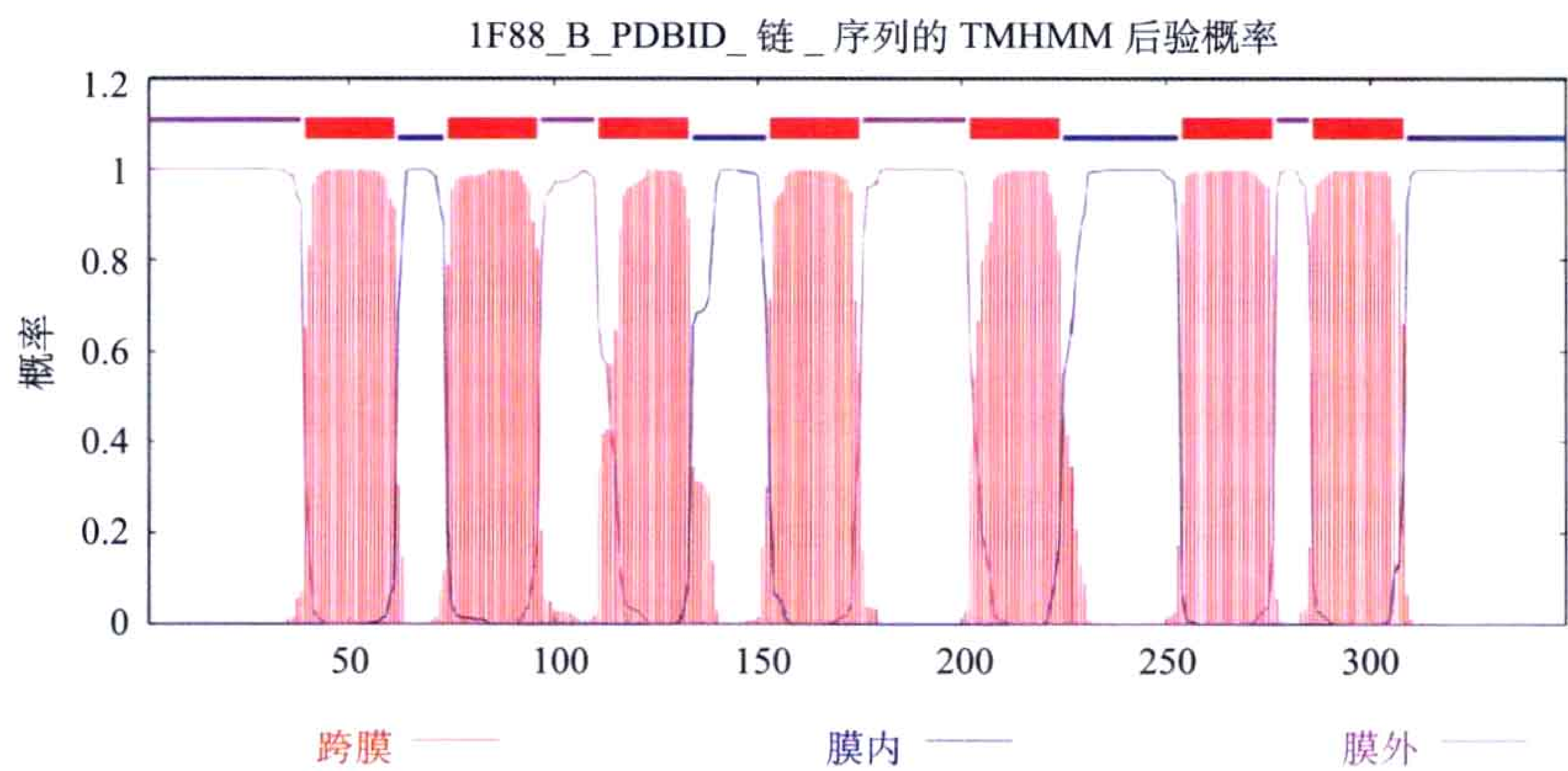


图 C.7 服务器 TMHMM 中的牛视网膜色素疏水图。

C.2.2 跨膜 β 桶状结构的预测

若干个跨膜的β桶已被识别。最早只是在许多的孔蛋白中找到这一类型的跨膜结构。后来发现这一现象可被重现，只是到目前为止仅在革兰氏阴性菌的外层膜被发现。桶状结构往往是由折叠束以反平行的方式组成，而折叠束的数目一般为偶数，在 8 ~ 22 之间。在一些情况下，这些 β 桶以孔的形式存在。其侧链当指向膜脂或孔时是相互交替的。那些与膜脂接触的残基通常是疏水的，而指向内部的残基则是亲水的。这与许

多其他 β 折叠很相似。利用已知的结构，在隐马尔可夫模型（hidden Markov model）的基础上，一些合理预测跨膜 β 桶的方法得以建立，但这些方法不能像预测跨膜螺旋那样准确。

C.2.3　无序片段的预测

大多数蛋白质都具有规则的构象，但是不同程度的灵活性对这些蛋白质行使功能来讲往往是非常重要的。许多蛋白质包含无序的片段。在有些蛋白中，这种无序片段不具备任何显著的功能；而在另一些蛋白中，这些无序片段则显得至关重要。这些无序片段当与其他分子相互作用的时候可以由无序变得有序，这一特性可以被用来改善结合的亲和力以及特异性。一些病毒外壳蛋白和核糖体蛋白具有这样的片段，当它们结合到其他蛋白质或核酸时才能变得有序的片段。另外还有一些蛋白质，表现为完全的无序状态，除非它们与其他大分子结合。无论是对功能的分析还是在结构的研究中，选择合适的片段以便对无序结构做出预测是非常重要的。

蛋白质中的无序区域有些时候也被称为“低复杂性”区域，其中氨基酸的分布不同于球蛋白中氨基酸的分布。这些区域中可能有非同寻常高频率出现的赖氨酸、谷氨酰胺、谷氨酸、甘氨酸及脯氨酸。对无序片段的预测可以通过神经网络来实现，该网络是被无序片段或整个蛋白质训练过的。这些程序具有一定的准确性。

C.3　蛋白质三级结构的模建

C.3.1　通过折叠来寻找同源模型——折叠识别

大量实验测定的结构使得利用一个适合的已知结构的蛋白质来进行模建（同源模建或比较模建）被广泛应用。这一方法的基础是尽管同源蛋白不具备显著的序列相似性，但是却拥有相似的结构。

对许多蛋白质来说，明确地找到一个已知结构的同源蛋白是不容易的。最有效建立同源性的方法是基于序列以及同一家族蛋白的序列概括图。这就是运用在 PSI-BLAST 中的方法。当这些方法不能够找到一个同源蛋白时，其他一些利用结构信息来辅助比较的程序可以被应用。这些程序是从所有已知折叠的数据库中选择最有可能的蛋白质折叠，通常这被称为折叠识别。

所有的折叠识别程序都利用大量已观察到的蛋白质折叠。主要的问题是要找到未知蛋白（目标蛋白）与已知折叠的正确比对。一些这样的折叠识别程序是利用动态编程的算法将已知折叠的蛋白质特性和序列与结构未知的序列进行比较，从而将已知的折

叠转化为计算机语言来编程。这些方法有些时候被称为剖面法。在这些方法中，序列被明确地用来与折叠进行比较。

利用这种方法进行的三维结构模建包括三个阶段。首先，正确的折叠类型必须被识别；其次，通过一些打分函数将具有未知结构的蛋白质序列与已知结构的蛋白质序列进行比对；最后，通过同源模建的技术将未知结构的蛋白质进行模建。

C.3.1.1 3D/1D- 剖面（3D/1D-profile）

在首先建立起来的折叠识别方法中，已知折叠的每一个位点都根据已观察到的二级结构和侧链环境进行了分类（3D/1D 剖面）。所得的线性剖面作为矩阵被编码，其中每一类氨基酸残基（或是间隙）都有一个分值，对应于已知折叠中在特定位置找到这一残基（或间隙）的可能性（图 C.8）。一个序列可以通过动态编程技术与这一剖面进行比对，所得比对分值的总和被用来衡量目标序列与该折叠的匹配程度。这一过程被称为反向折叠，是由于任何一个序列都可以被用来测试它与某一折叠的兼容性。这一过程可被用来检测晶体学实验中得到的不正确的模型。当氨基酸残基不能正确地与实验电子密度图相吻合，因此也不能与它“自身”的剖面相匹配时，因此所得分值就会很低，如当极性残基指向疏水中心时。

Pos	Class	A	C	D	E	F	G	…	R	S	T	V	W	Y	Gap open	Gap ext.
1	E	12	–46	22	3	–190	113		–32	32	12	–91	–214	–94	2	0.02
2	B_2	–66	–5	–128	–135	105	–166		–80	–117	–76	60	102	112	2	0.02
3	E α	46	–44	44	59	–220	68		–34	15	–17	–110	–135	–210	200	200
4	P_2 α	6	–93	28	56	–143	–50		50	–18	–5	–48	–114	–79	200	200
5	E α	46	–44	44	59	–220	68		–34	15	–17	–110	–135	–210	200	200
6	P_2 α	6	–93	28	56	–143	–50		50	–18	–5	–48	–114	–79	200	200
7	B_2 α	–69	–10	–162	–71	90	–149		6	–147	–150	68	50	85	200	200
8	E α	46	–44	44	59	–220	68		–34	15	–17	–110	–135	–210	200	200
9	P_2 α	6	–93	28	56	–143	–50		50	–18	–5	–48	–114	–79	200	200
10	B_1 α	–66	–73	–197	–174	132	–253		–167	–273	–129	66	100	18	200	200

图 C.8 蛋白质中一个短片段的 3D/1D 剖面。折叠中每一位点根据其所处的环境，被归为 6 类中的一类（以极性增长为序：B_1、B_2、B_3 为包埋的；P_1、P_2 为部分包埋的；E 为暴露的）并且每一类又根据其二级结构被分为三类。每个氨基酸或间隙在这 18 种类型中都有一个分值。图中所示的是一些氨基酸的分值。疏水性残基通常是包埋的（B），反映在 B 类环境中较高的分值。暴露的疏水性残基和包埋的极性残基均在其相应的分类中给出负的分值。摘自 Bowie JU et al. (1991) Science **253:**164-170。

C.3.1.2 剖面法（Profile method）

上述的 3D 剖面可以作为位置特异的打分矩阵，通过动态的程序算法来进行比对。一些成功的折叠识别方法也通过这一算法将已知的折叠转化为条目，然后将未知的序

列与这些条目进行比对。TOPITS 是一个实例。在这一方法中，多条与未知序列同源的序列被比对，比对结果被用于二级结构的预测以及通过 PHD 神经网络的方法来估计表面易接近度（图 C.9）。这一方法中，序列被转化为 1D 剖面，其中每一位点包含氨基酸残基以及它所预测的二级结构和表面接近度。结构数据库中所有的已知结构则以在每一位点上具有相同氨基酸残基和已知二级结构及其表面接近度的同种类型已观测剖面的形式加以储存，然后通过动态程序算法来进行比对。对每一位置的打分是一个适合的权重和，该权重介于普通的取代矩阵分值（如 BLOSUM62）和简单的结构剖面矩阵之间。另外，还存在间隙权重。

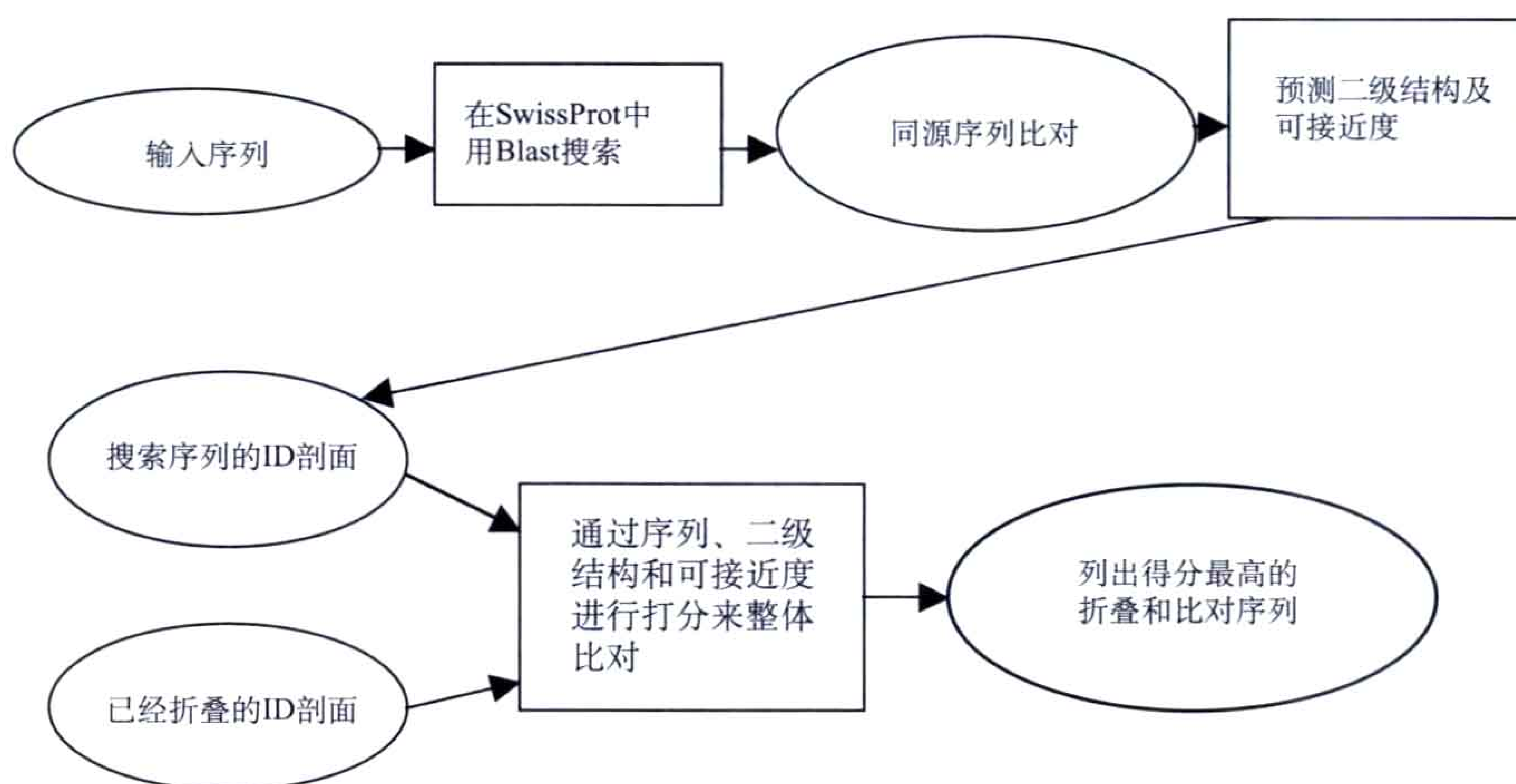


图 C.9 3D/1D 法（TOPITS）的流程图。第一步与用 PHD 法预测二级结构相同。

这些方法是简单并且自动化的，但是当将结构转化为 1D 剖面时，显然有关在已知结构中残基的空间相互作用的信息是缺失的。

C.3.1.3 穿针引线法（threading method）

在穿针引线法中，序列被“粘贴”到已知的折叠上，序列与折叠的兼容性则通过对这一模型的分值或能量的计算来进行评估。打分或势能函数可以通过不同的方式进行构建，所得分值应该与具有相似长度的不同序列（不相关序列）的打分所得分值进行比较，以便测定任意高分的显著性。

一些步骤是基于蛋白质中不同残基对之间相互作用的概率来实现的。这些概率被用来构造一个残基间成对势能的矩阵（图 C.10）。每一折叠则通过残基间的相互接触加以分析，结果产生一个相互间距离在一定范围内的残基对表格。接下来，未知蛋白的序列被缠绕到每一折叠上，对于与该折叠的每一次比对，其总势能则可通过成对相互作用势能矩阵来进行计算。这一方法及与其相类似方法的成败取决于这些势能或打分矩阵，而这些矩阵应该遵循这样的事实，即疏水的侧链大都集中在疏水中心，而亲水残基则倾向于暴露在外面。

距离	A–A	R–A	R–R	N–A	D–R	D–N	D–D	I–E	L–I	T–S
0–5	0.230	0.237	–0.145	0.159	–0.527	–0.534	0.539	0.385	–0.223	–0.430
5–6	0.350	0.338	0.024	–0.170	–0.618	–0.481	0.128	0.537	–0.518	0.033
6–7	0.223	0.101	0.215	0.042	–0.609	–0.403	–0.006	0.464	–0.702	–0.169
7–8	0.284	0.266	0.249	0.010	–0.243	–0.144	0.036	0.158	–0.347	–0.056
8–9	0.020	0.068	0.195	–0.071	–0.266	–0.321	0.021	–0.019	–0.071	–0.020
9–10	0.017	0.050	–0.035	–0.150	0.130	–0.075	0.033	0.122	–0.030	–0.050

图 C.10 几个 6×21×21（20 个氨基酸加 1 的多肽组）将序列缠绕在折叠上的接触势能。距离是位于由骨架坐标计算得到的侧链的标准位置之间的距离。如果两个丙氨酸残基在 10Å 以内，A-A 列中所显示的势能便被加入总的分值中。负的势能值代表一个有利的接触。势能是由蛋白质中所观察到的接触频率计算得到。摘自 Bryant SH, Lawrence CE.(1993)*Protein Struct Funct Gen* **16:** 92-112。

相对于那些具有代表性的折叠来说，大多数序列具有间隙和删截，因此打分就不得不通过与该折叠的多种可能的比对来进行评估。在这种情况下，不太可能运用动态编程的方法，这是由于与不同区域的比对不是相互独立的。为了避免计算与折叠所有可能比对的分值，穿针引线法采用了一些简化的办法，使得打分在较合理的时间内得以计算。这仍然是该方法在发展中的一个重要的局限。

C.3.1.4 组合方法

为了改进结果，可以将穿针引线法和序列剖面结合起来。基于这两者基础上建立的方法可以更好地预测并且进行折叠与序列的比对。其原因大概是由于这两者是互补的。序列剖面可以探测较弱的序列相似性，但是这些方法不似穿针引线法的势能，对疏水核心那么敏感。基于不同的算法，可以将许多的预测服务器组合起来，建立后服务器。用不同方法预测的折叠要比仅用单一步骤预测所得折叠的可信度高得多。

C.3.1.5 CASP 和疑难案例的成功率

最好的检测预测方法是将它们用在那些结构尚未被解析出的蛋白质上。这些测试是蛋白质结构预测技术评估（Critical Assessment of Techniques for Protein Structure Prediction，CASP）会议的关注点。在会议上，参与模建和方法发展的科学家被邀请来对一些结构即将通过实验方法被解析的蛋白质进行模建。这些模建将被提交并且在当实验测定结构出来时与其作比较。这类两年一次的会议对这一领域的发展非常重要。

近几年，折叠识别方法的成功率已经有所提高。在序列相似性很难确定时，有相当一部分序列的正确折叠已经被得到。可对未知蛋白指认其折叠从而提供一个可能的功能，这些程序是非常有用的。然而，即便这些步骤可以成功地预测出正确的折叠，其比对也往往是不正确的。由于在序列同源性较低的情况下进行同源模建是困难的，基于折叠识别方法和不准确的序列比对建立起来的蛋白质三级结构模型不是非常可靠的。

在折叠识别试验中，正确的折叠通常与几个能量及打分相似的其他折叠一起被识别出来。因此许多努力被用来优化打分函数的参数，以便从这些折叠中识别出正确的折叠来。然而，也存在所有情况下势能方程都不会在模型间被正确识别出来的可能。其原因是即便正确折叠的自由能已知，它可能并非显著低于其他一些折叠。由于这是一个折叠过程的结果，天然结构的能量并不一定低于其他折叠或是非折叠的能量，折叠的动力学对其最终构象可能有重要的影响。

C.3.2　基于已知结构的模建

基于已知结构的蛋白质三维结构的模建具有不同的名称：同源模建、比较模建或模板依赖的模建等。该方法的第一步是确定一个或几个与待模建（目标）蛋白质有足够相似性的合适的起始模型（模板或者母体）；第二步包括将未知蛋白与已知结构的序列进行比对；第三步是将蛋白质进行模建。

C.3.2.1　寻找合适的模板

显然，同源模建是以目标蛋白与模板蛋白同源并且具有相同的折叠这一假设的正确性为基础的。可以通过 BLAST 将目标序列与 PDB 文库中的所有序列进行比对。如果存在多个可能的模板，那么与未知蛋白具有最高相似性的序列是最好的选择，因为其结构与目标蛋白具有最大限度的相似性。然而，通过一种以上的模板进行模建也会提供很多有用的信息。当没有明确已知结构的同源蛋白来进行模建时，也可用其他办法来寻找具有远亲关系的模板，如穿针引线法（见 C.3.1 节）。

C.3.2.2　模板和目标序列的比对

由于序列比对中的错误在模建过程中不容易被纠正，所以比对步骤非常关键。如果序列的相似性高并且序列中没有插入与删除，则比对很容易进行。如果序列并不十分相似，就需要正确地安置删除与插入的位点。一些通过自动比对步骤引入的错误可以运用知识人为地调整蛋白质删除和插入的位点。在这一阶段，通过比对众多的同源序列来确定蛋白质家族的保守和可变区域将是非常有效的。然而，对蛋白质的环状区域或者末端进行正确比对往往是非常困难的。

C.3.2.3　目标结构的模建

第三步需要作出判断以确定已知结构中哪些部分应作为起始模型，哪些部分在目标结构中被预期难以独立进行模建。通常模板蛋白中的二级结构元素可用作起始模型，但这取决于序列的相似性，一些环状区域也可以被包括进来。用已观察到的环状区域的数据库可以对未知构象的环状区域进行模建。

在保守的区域中如果有可能应尽量保留侧链的构象，而那些在目标和模板序列中不同的残基则可以被取代。由于空间位阻的原因，氨基酸的侧链往往仅限于有限的几种构象，即旋转异构体（见 2.1 节）。每一种氨基酸残基都有其自身偏好的构象，因此侧链修饰过的残基构象应该从该氨基酸偏好的旋转异构体中产生。

C.3.2.4　同源模型的质量

同源模建可以产生较为准确的模型，特别是当模板与目标蛋白具有高度的序列相似性时。对蛋白质核心部分模建产生的模型可以期望达到较好的质量水平，但是对环状区域部分来说则较差。最新的一些方法改进有时至少能够在局部范围内得到比模板更接近真实结构的模型。对生物学家来说，这些模型因此可以作为基础被用来检测对功能做出的假设。当序列的相似性较低时（小于 30%），基于序列的同源模建很可能是不完全正确的。即便折叠已被正确识别，进行序列比对的困难使得不仅在环状区域内，甚至在二级结构单元内的序列也不能够被恰当地归属。另外，当起始模型与真实结构的构象差距较大时，模建也不容易得到正确的构象。由于蛋白质的活性位点和一些对功能较为重要的部分相对于其他区域通常更为保守，对这些部分的同源模建对于预测可被检测的功能特性是十分有效的。

C.3.2.5　同源模建的程序

SwissModel 是一个通过上述步骤来进行同源模建的服务器。首先，在已知结构的数据库中通过 BLAST 来找到与搜寻序列有显著序列相似性的、适合的、已知结构的、蛋白质；接着进行序列比对。在这两个阶段，用户可以与服务器相互交流以选择模板结构或者调整序列比对。一个包含有大量模型的储存库也可以通过这一办法产生出来。同样，ModBase 和 RIKEN Famsbase 也是较大的同源模建的数据库。

一些最为成功的自动同源模建服务器是建立在较为复杂的步骤之上的，这些步骤是基于克服不正确比对以及对表面特性较差的模建之上的。

C.3.3　没有同源结构的模建：*ab* 从头算法模建

在没有相似结构的特殊模板的情况下进行三维结构的模建已经有很长的历史了。这些方法被称为 *ab* 从头算法、新折叠预测方法或是自由模建方法。早期的方法通过简化描述的氨基酸残基来计算构象的最小能级。这些步骤试图去模建蛋白质折叠的通路，因此也被称为蛋白质折叠的 *ab* 从头算法。由于最初的方法遇到了来自潜在势能方程和计算能力的限制，这些方法与折叠识别的方法相比，成功的案例有限。通过计算机来进行蛋白质折叠仍然是一个非常有活力的研究领域。分子动力学的方法可以被用来在溶液中模仿蛋白质分子的轨迹，但是这些计算只能用于那些比起实际折叠短得

多的时间尺度上。

一些新的方法已显示出在预测新结构时的用处，Rosette 方法就是其中一个例子。这个方法基于这样一个重要的假设：蛋白质中一个短片段的折叠与其他蛋白质中具有相似序列的局部折叠相似。由 3 个和 9 个残基组成的折叠数据库可以被用来预测未知结构的折叠。在这个数据库中，包含有相同序列的不同折叠的例子。这些片段可以以很多方式组合起来，并且其偏好于成对 β 束和包埋疏水残基的分值被计算出来。由这一方法产生出大量的模型，这些模型以相似的折叠被归类。其中最大的一类折叠被选择作为预测的折叠。尽管优于许多早期的预测方法，这个方法及其他一些基于片段的步骤只能对一小部分序列找到其正确的折叠，以这些方法获得的模型并不是十分准确。

延伸阅读

Bonneau R, Strauss C, Rohl C, Chivian D, Bradley P, Malmstr ömL, Robertson T, Baker D. (2002) De novo prediction of three-dimensional structures for major protein families. *J Mol Biol* **322**: 65.

Bowie JU, Luthy R, Eisenberg D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**: 164-170.

Rost B, Sander C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**: 55-72.

von Heijne G. (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**: 487-494.

数据库和服务器的链接

CASP: http://predictioncenter.org/

ModBase: http://modbase.compbio.ucsf.edu/modbase-cgi/search_form.cgi

SwissModel: http://swissmodel.expasy.org/

SwissModel Repository: http://swissmodel.expasy.org/repository/

PredictProtein (Server for several prediction programs): http://www.predictprotein.org/

3D-PSSM, fold recognition and other prediction programs: http://www.sbg.bio.ic.ac.uk/

Membrane-spanning barrels: http://cubic.bioc.columbia.edu/services/proftmb/

附录 D
蛋白质功能的确认

D.1　通过序列相似性鉴定功能

D1.1　序列相似性和同源性

超量的基因组序列信息要求我们建立起各种方法来发现未知基因编码的蛋白质的可能功能。蛋白质功能最主要的注释方法是通过序列比对，找出与已知功能蛋白的同源性来设定未知蛋白的功能。两个序列之间相似性用序列比对分值（alignment score）来衡量，结合适当的统计学处理，比对分值就可以给出一个可能的功能判断。该注释方法的前提是假定真正不相关的序列没有系统相似性，然而实际情况则不尽然。某些蛋白质含有大部分由几个氨基酸残基重复而成的区域（即所谓的"低复杂性区域"）。一般来说，蛋白质含有非随机序列是源于它需要形成堆积良好的疏水核心所致。

如在第二章和第十六章中所述，有些蛋白质虽然具有相同的折叠类型，但并无证据显示其起源相同。这种相似性是由结构收敛（structural convergence）造成的。结构收敛是指不同起源的蛋白质各自独立进化而形成相同的结构折叠类型。理论上说，序列相似性也可能是源于功能上的限制由不同起源的蛋白质进化趋同所致。尽管在蛋白质中存在少量氨基酸残基具有收敛适应性变化的例子，但是全长蛋白质或结构域的序列收敛性并未被发现。因此，结构域的显著相似性通常可以用来证明蛋白质的同源性。

虽然决定两个蛋白质是否同源的最好方法是对序列比对分值进行适当的统计学分析，但通常用序列一致性程度来衡量蛋白质同源性更为方便。多序列和结构比对分析的结果显示，若两条序列具有 35% 的一致性（两条序列中一致的氨基酸残基数 / 两条序列中较短序列的氨基酸残基总数），则可以视为两条序列同源。对于较长的序列，可以使用更低一些的一致性百分比阈值（图 D.1）。在阈值线以下的灰色区域（有时也标记为模糊区）里可以找到真正的同源蛋白，但是不能断定相似性是否来源于同源性。

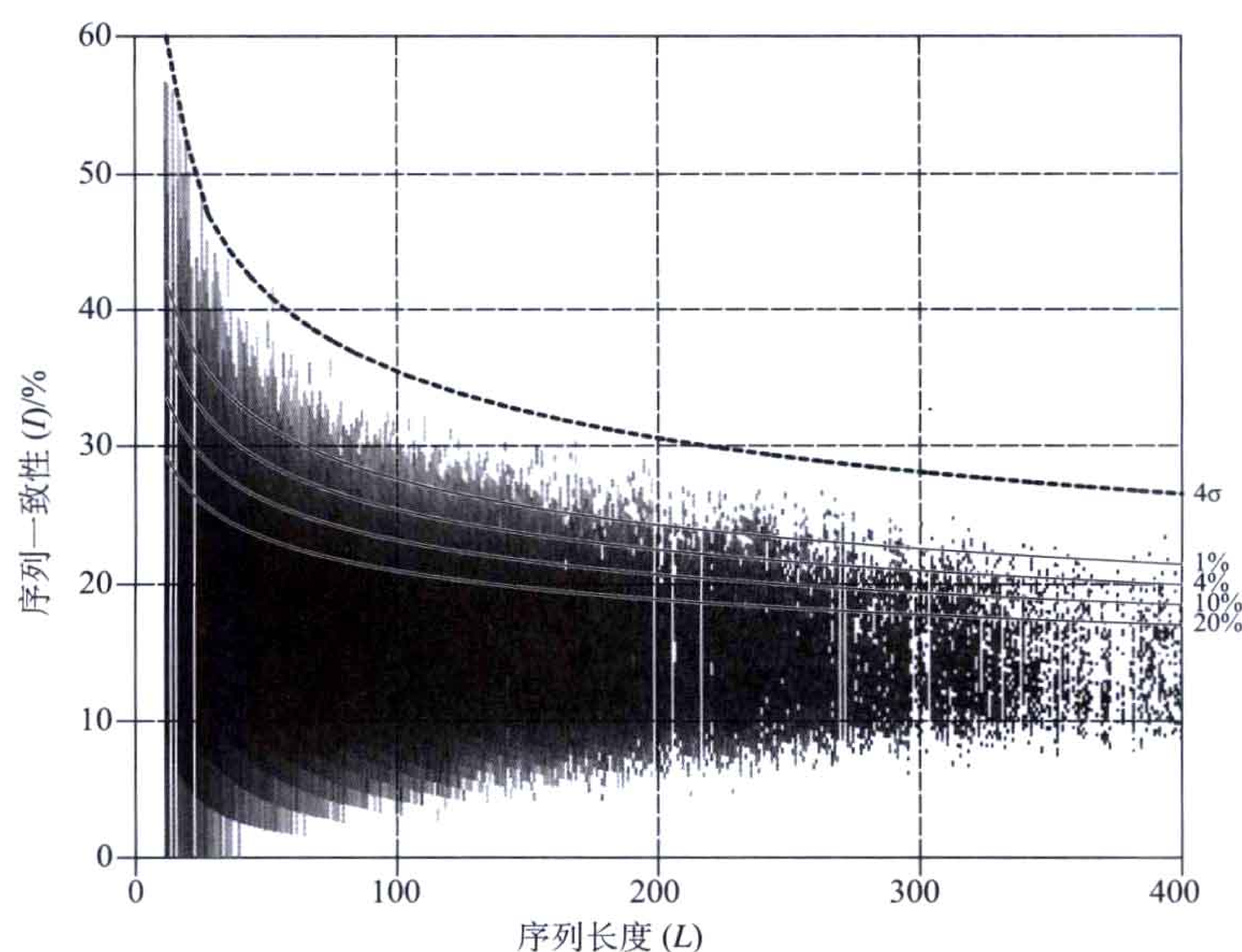


图 D.1　通过对 130 万个不相关的不同长度的蛋白质序列比对得出的蛋白质一致性与序列长度的关系图。序列比对采用 Needleman-Wunsch 算法进行，序列末端的 gap 不罚分。不同长度的蛋白质序列同源性阈值在图中用曲线表示 [图片来自于 Batalous，略有修改。经 Abagyan RA 和 Batalov S (1997)（*J Mol Biol* **273:** 355-368）授权再印]。

如果序列一致性较低，则需要更为复杂精密的方法来测定序列的相似度，如量变曲线（profiles）或隐马尔可夫模型（hidden Markov model）等方法。

D.1.2　保守序列样式

在很多情况中，同源蛋白的保守氨基酸残基显示出特定的样式。由于功能重要的氨基酸残基倾向于保守，因此这种保守样式可以被试着用于鉴定未知蛋白，在缺少总体序列相似性的情况下将蛋白质与功能相联系。例如，结合血红素的细胞色素 c 分子都含有保守序列 CXXCH，X 可以是任意的氨基酸残基；两个半胱氨酸与血红素共价结合；组氨酸是血红素中铁原子的一个配体。在另外一些蛋白质中，保守的催化残基或参与形成二硫键的半胱氨酸残基也可以用于蛋白质的识别。

有些程序已经用于识别上述这些或其他保守模式，ProSite 数据库就已经定义了很多的保守样式并用于鉴定相关蛋白质。然而最主要的问题是建立特异性的保守样式比较困难。例如，在 SwissProt 中搜索 CXXCH 序列，命中的许多结果却并不是细胞色素。这些假阳性的搜索结果是不相关的蛋白质，它们不结合血红素，并且构造也不同，因此 CXXCH 不能作为用于寻找细胞色素的样式。

胰岛素含有一个特征序列样式 C-C-{P}-{P}-x-C-[STDNEKPI]-x(3)-[LIVMFS] - x(3)-C，大括号 {} 中的残基表示不允许的氨基酸残基，中括号 [] 中的为允许的序列段。这个样式表示有 4 个保守的半胱氨酸残基形成 2 个二硫键，使用这个保守样式在

SwissProt 中可以搜索到大约 200 条胰岛素链，但是丢失了 10 条（胰岛素）链，而且命中结果中有 3 个是错误的。

球蛋白家族的保守序列非常短（详见第十六章），搜索短的保守序列会产生很多的假阳性结果。使用涵盖了一个完整序列和在某个特定位置编码特定氨基酸残基信息的 sequence profile 代替上述短的保守序列进行搜索，则可以更精确地鉴定球蛋白。将该序列与特定的保守序列文件比对并计算出比对分值，如果最高比对分值大于阈值，则将该序列归为命中序列。在 ProSite 中，采用这种搜索方式已经在 SwissProt 中鉴定了 739 个球蛋白，其中只有一个假阳性结果。

D.1.3　同源蛋白是否具有相同的功能?

大多数情况下，蛋白质之间的显著序列相似性意味着它们具有相同或类似的功能，但也存在例外。最为极端的例子是“兼职”，同一个蛋白质在一个物种中具有两种或两种以上的功能。这种蛋白质的例子很多，而且其不同的功能可以完全不相关。这种现象的著名例子是几个眼睛水晶体中的晶状体蛋白 crystallin，它们具有代谢酶的特性，同时在晶状体中大量表达并充满整个晶状体。在这个例子中，不同类型的细胞利用蛋白质的不同功能。在另外一些例子中，蛋白质在细胞内和细胞外具有不同的功能。

同源蛋白可能是直系同源，指同一个基因在物种进化中发生改变；也可能是旁系同源，指基因复制使得在同一物种中产生了编码同一蛋白质的基因的两个拷贝，而这两个拷贝经过基因突变进化成为两个不同的蛋白质。一般认为直系同源的蛋白质具有相似的功能，但是旁系同源则可能具有多种不同的功能。旁系同源的蛋白质即使具有很高的序列相似性，也不一定就意味着其功能非常相似。例如，鸡溶菌酶和山羊 α- 乳清蛋白具有高达 45% 的序列一致性，但是这两个蛋白质的功能非常不同。然而，大多数的旁系同源蛋白在同一物种中被用来行使与其起源蛋白相关的功能。

D.1.4　序列比对数据库和序列保守模式

由位于英国的欧洲生物信息学会（European Bioinformatics Institute）维护的 InterPro 数据库正尝试将不同数据库中蛋白质家族相关信息整合在一起。Pfam 数据库（英国）含有大量的结构域家族的比对序列，值得注意的是在 Pfam 数据库中，蛋白质是按照结构域分类的，所以多结构域的蛋白质可以在多个蛋白质家族中找到。序列比对是用隐马尔可夫模型进行的，每个蛋白质家族隐马尔可夫模型（hidden Markov model，HMM）是基于 Clustal 软件对一定数量的序列比对并结合手动调整 [称为“种子比对”（seed alignment）] 而建立的，然后再通过 HMM 在非冗余蛋白序列数据库（SwissProt+TrEMBL）中进行搜索，将其他蛋白质加入蛋白质家族中。在 Pfam 数据库中，蛋白质的结构域组织关系用图表显示，搜索序列也可以通过同样的方式进行分析。数

据库中存在两种类型的蛋白质家族：A 型，是明确定义并记录在案的；B 型，是没有记录在案的同源序列。Pfam 数据库包含近 9000 个蛋白质家族。

很多的数据库和服务器拥有蛋白质保守序列特征信息。由瑞士生物信息学会（Swiss Institute of Bioinformatics）维护的 ProSite 数据库包含将近 1500 个可以用于鉴定蛋白质功能的保守序列样式信息。该数据库中拥有一个涵盖每个保守序列样式的条目列表（带链接）。阳性、假阳性和假阴性的结果都有标识。另外，该数据库还具有对保守序列样式以及蛋白质家族的说明。

D.2 基于结构的功能预测

D.2.1 结构基因组学

虽然确定蛋白质功能最主要的方法是寻找序列相似性，但是在很多情况下无法找到与所研究蛋白质序列相似的已知功能的蛋白质。在不存在显著序列同源性的情况下，另一种可以鉴定蛋白质功能的方法是解析蛋白质的三维结构，然后根据蛋白质的折叠类型预测其可能的功能。虽然一般来说蛋白质折叠类型与功能没有必然联系，但也有些折叠方式与一个或多个活性相关。

在利用结构基因组学研究蛋白质功能的领域中，最早的系统性工作是古细菌中一些蛋白质的 NMR 和 X 射线晶体学研究。该研究选取了功能完全未知的蛋白质为靶蛋白。有 1/10 的与已知蛋白质没有显著序列同源性的蛋白质是具有已知拓扑结构的，其中部分蛋白质可以根据结构相似性推测其功能。另一些例子中，结构解析后发现晶体中蛋白质结合有特定的配体，这也有助于进一步确定蛋白质的功能（图 D.2）。当然，也存在结构相似性只能给出一个非常粗糙的蛋白质功能，甚至不能提供任何功能推测线索的情况。

D.2.2 利用模板的功能预测

若蛋白质的三维结构已知，则可以根据局部结构相似性推测蛋白质的功能。酶活性位点的三维结构模板、DNA 结合和配体结合表面，以及由目的蛋白自身定义的反向模板可以用来寻找目的蛋白与功能已知蛋白之间的结构相似性。描述活性位点的模板含有 2~5 个氨基酸残基，而另一些模板含有 3 个氨基酸残基。ProFunc 服务器（EBI, UK）利用这种方法，将模板方法、序列比对以及整体结构比较相结合，用来预测未知蛋白的功能。

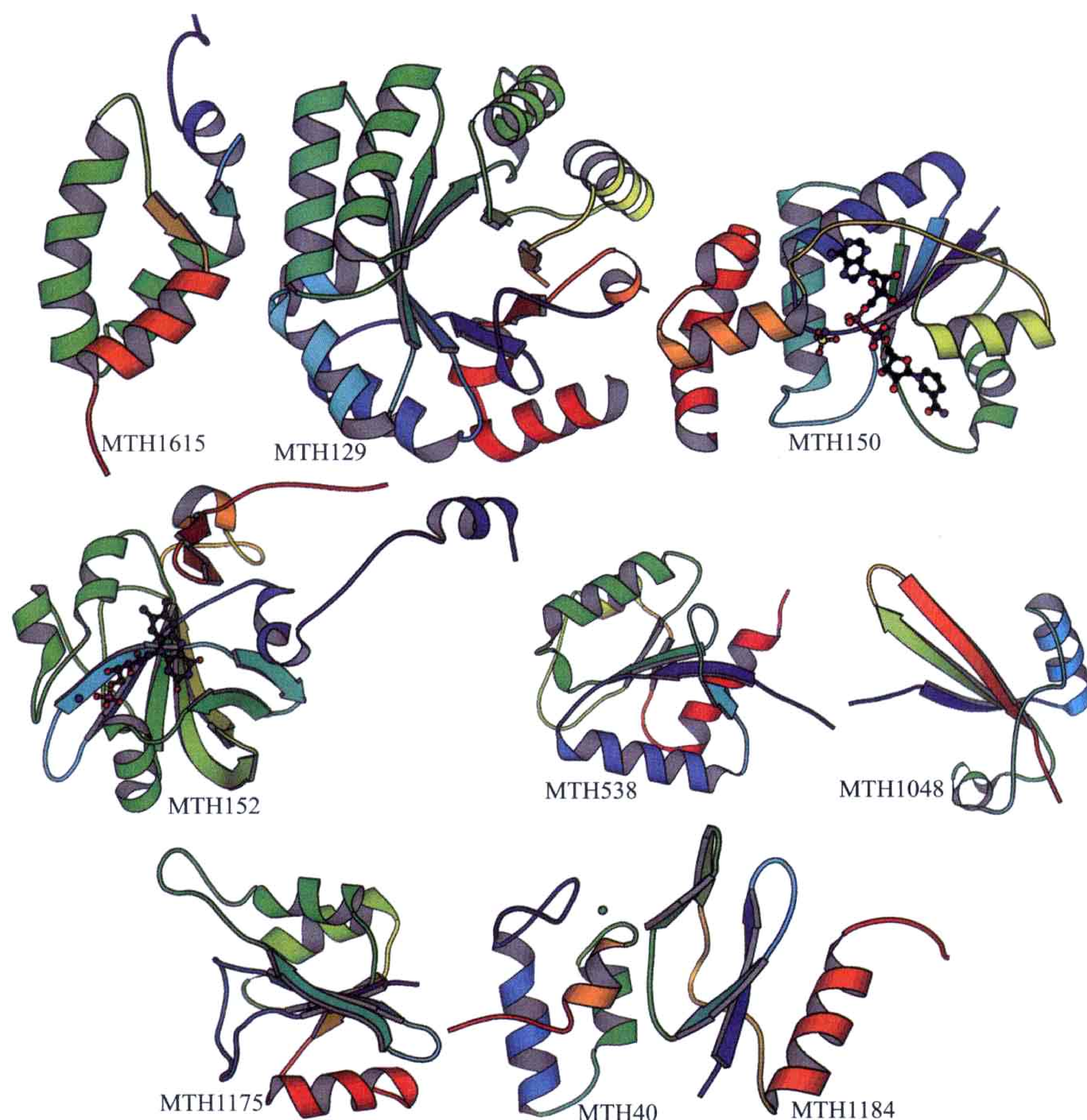


图 D.2　古细菌嗜热自养甲烷杆菌（*Methanobacterium thermoautotrophicum*）的 9 个蛋白质的结构模型，这些蛋白质在其结构被解析时与已知结构的蛋白质没有序列相似性。在晶体中，MTH150 和 MTH152 分别结合 NAD^+ 和 FMN，加上与已知功能的酶的结构相似性，可以推测出蛋白的功能（MTH150 被确定为腺苷酰基转移酶）。MTH129 是 oritidine 单磷酸脱羧酶的同源蛋白，由此可推测该蛋白质的功能，但其结构未知。根据结构可以推测 MTH1615 的功能可能为 DNA 结合 / 转录因子。MTH40 和 MTH1048 是 RNA 聚合酶Ⅱ的亚基。还有些蛋白质与已知功能的蛋白质结构类似，但是缺少重要的氨基酸残基，因此不能更特异性地鉴定其功能。

D.3　根据基因组比对进行功能鉴定

未知基因的功能也可以通过间接方法获得。了解两个蛋白质是否参与同一个代谢途径的一种方法是看这两个蛋白质在其他基因组中的同源蛋白是否融合成一个基因。在一个物种中某个基因的融合强烈地预示着融合基因编码的两个蛋白质在另一个物种中生理上的相关性或参与催化相继的反应。在细菌基因组中，已发现数百个蛋白质对在其他物种中以融合蛋白形式存在。

另一个鉴定蛋白质功能的方法是分析系统进化图，现在有很多此类基因组背景方法（genomic-context method）。在基因组中同时存在或不存在的蛋白质对或蛋白质组很有可能在生理上或功能上相关，保守的基因序列或共享的调节元件也预示着蛋白质功能的相关性。

D.4　蛋白质与蛋白质相互作用

虽然已经发现了很多蛋白质复合体并且鉴定了特性，但是通常还是倾向于将蛋白质视为独立行使功能的单独实体。然而，越来越多的证据显示，在细胞中很大一部分蛋白质形成复合体。既然全基因组都已被鉴定，研究蛋白质在细胞中是如何组织的就尤为重要。绘制蛋白质与蛋白质在细胞中的相互作用图谱是一项工程浩大的工作。经过生物化学家们多年的研究，现已经积累了大量的蛋白质间相互作用的数据，而基因组测序工作中获得的未知功能蛋白数量的大幅增长更增加了蛋白质相互作用信息的有效分析和储存的工作。

通过实验获得此类信息的一种方法是双杂交实验，可以测试蛋白质组合之间的较强的相互作用。这种方法需要将蛋白质文库与"诱饵"蛋白在细胞中共表达，利用遗传学方法，只有表达了与诱饵蛋白相互作用的蛋白质的细胞才会生长。这种方法已经被用于酵母和其他一些物种中寻找蛋白质与蛋白质的相互作用。

另一种方法是通过合适的纯化标签直接从细胞中提纯蛋白复合体，然后利用质谱的方法鉴定复合体中的蛋白质成分。

实验性的方法并不十分可靠，它要求蛋白质的相互作用结合能力达到一定水平。目前也有建立计算方法来有效预测蛋白质间相互作用方面的尝试性工作。但是由于相互作用表面通常与蛋白质其他部分的表面没有什么不同，使得发展计算性的方法比较困难。相比之下，鉴定结合洞（cavity）和沟（cleft）则较为简单，而这些结合洞和沟一般用来结合小分子。

在 BIND 数据库中可以获得蛋白质相互作用的信息。该数据库由加拿大多伦多西奈山医院维护，目前有 117 000 个条目。该数据库中的信息不限于蛋白质与蛋白质的相互作用，也有蛋白质与核酸和小分子配体的相互作用信息。

延伸阅读

Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I, Kozlov G, Maxwell KL, Wu N, McIntosh LP, Gehring K, Kennedy

MA, Davidson AR, Pai EF, Gerstein M, Edwards AM, Arrowsmith CH. (2000) Structural proteomics of an archaeon. *Nature Struct Biol* **7**: 903-909.

Laskowski RA, Watson JD, Thornton JM. (2005) Protein function prediction using local 3D templates. *J Mol Biol* **351**: 614-626.

数据库和服务器链接

InterPro: http://www.ebi.ac.uk/interpro/
Pfam: http://www.sanger.ac.uk/Software/Pfam/
ProSite: http://expasy.org/prosite/
ProFunc: http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/

附录 E

蛋白质的修饰

在细胞中，蛋白质可以发生多种方式的翻译后修饰，这些修饰反应可以是自发的也可以是酶催化的，在有些情况下是由多种酶催化的反应。很多修饰在功能上是必需的，会导致蛋白质的活化或失活。

酶催化反应有时需要辅因子或辅酶的协助，这些辅因子或辅酶可以通过共价或非共价的方式与酶结合（见 2.2.4.3 节）。这些辅助因子是单独合成的。也有些被编码氨基酸残基的翻译后修饰会在酶的活性位点形成新的结构。 很多这类修饰作用首先是通过结构研究而被鉴定出来的，然后由化学方法加以证实。结构生物学家应该密切关注那些可能来源于蛋白质修饰的不寻常特征，而这些新型修饰结构的生物合成是大家关注的热点。在此，我们无法对蛋白质翻译后修饰问题的复杂程度进行全部讨论，要了解更多信息，请参考本章后附的参考文献。

地址标签标记是修饰的一种类型，它控制蛋白质未来的命运及目的地，如泛素化介导的蛋白质降解（详见第九章）。蛋白质修饰也会引起损伤，如被氧化或被亚硝基化。本附录将选择性地介绍蛋白质修饰的几个例子。

E.1 氧化作用和羰基化作用

活性氧（ROS）会导致生物分子的氧化。细胞中主要的活性氧有过氧化氢（H_2O_2）、羟基自由基（•OH）和超氧自由基（O_2^-•）几种形式。在真核生物中，它们主要产生于线粒体，但也会产生于叶绿体和过氧化物酶体中。活性氧可以是自由基也可以是非自由基氧化剂。少量的活性氧对细胞行使正常功能是必需的。正常情况下，活性氧在细胞中存在一个产生和消去的平衡。抗氧化防御系统包括超氧化物歧化酶、过氧化氢酶和过氧化物酶。

半胱氨酸氧化形成二硫键属于自发的修饰，很多酶都存在特异性的二硫键形成与还原过程。半胱氨酸可以被进一步氧化成为次磺酸（C-SOH）、亚磺酸（C-SO_2H）或磺酸（C-SO_3H）（图 E.1）。甲硫氨酸可以被氧化成亚砜或砜类。硫的氧化是可逆的。

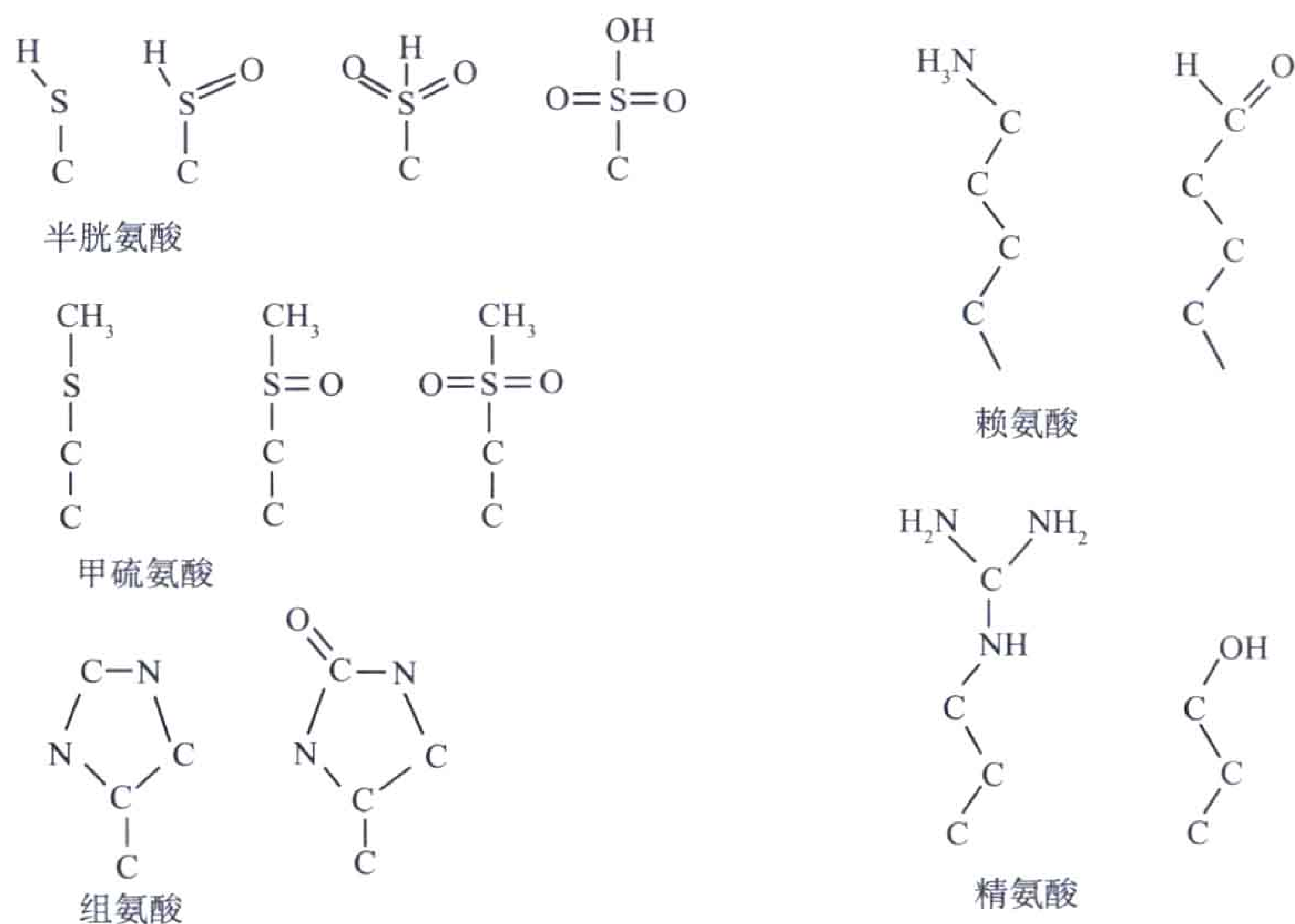


图 E.1 氨基酸残基侧链氧化修饰的例子。

除了含硫蛋白，其他许多氨基酸侧链也可以被氧化，其中最容易被氧化的是组氨酸和赖氨酸（图 E.1）。这些修饰经常是不可逆的。这种类型的自发氧化被称为羰基化。例如，脯氨酸可以被羰基化形成 2- 吡咯烷酮，精氨酸被羰基化形成谷氨酸半醛，组氨酸被羰基化形成 2- 氧化组氨酸，赖氨酸被羰基化形成氨基己二酸半醛等（图 E.1）。羰基化反应很难诱发，但是它的发生是不可逆的。氨基酸残基的羰基化衍生物是通过金属直接氧化进攻形成的。很多人类疾病与蛋白质的羰基化作用相关。在大肠杆菌中，诸如分子伴侣 Hsp70 和 Hsp60（详见第九章）、延伸因子（详见第八章）或丙酮酸激酶都是羰基化反应的靶标蛋白。

赖氨酸羰基化反应的详细研究发现，只有很少数特异性的氨基酸残基被修饰。被修饰的赖氨酸位于一个或几个酸性氨基酸侧链附近，而这些酸性氨基酸的侧链很有可能参与和赖氨酸形成盐桥，或为金属提供暂时的结合位点。

在细胞中，与野生型蛋白相比，被羰基化修饰的蛋白质通常比较不稳定，容易被水解，也更容易发生聚集。现已发现，蛋白质羰基化作用的增加与组织的老化存在一定的关联性。阿尔兹海默症患者脑中淀粉样老年斑的主要成分 β- 淀粉样蛋白(amyloid-β, Aβ），已被发现是通过产生活性氧而增加氧化压力。

E.2 羟基化作用

脯氨酸和赖氨酸残基在翻译后可以发生酶催化的羟基化反应。由脯氨酸羟基化酶催化的脯氨酸残基羟基化反应通常发生在 Cγ 上（图 E.2）。羟脯氨酸（三个字母缩写

为 Hyp）是胶原蛋白非常重要的组分（详见第十三章），存在于胶原蛋白的 Gly-X-Hyp 重复序列中，X 通常为脯氨酸。羟赖氨酸也存在于胶原蛋白中。

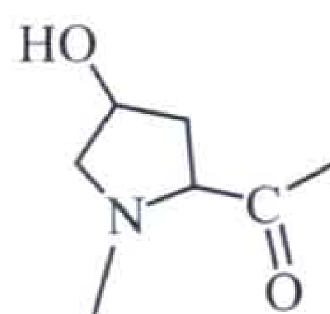


图 E.2　羟脯氨酸（Hyp），胶原蛋白的重要组分，存在于 Gly-X-Hyp 重复序列中。

E.3　脱氨基化作用

很多蛋白质修饰是由于增加了官能团。另一种更基础的修饰是天冬酰胺和谷氨酰胺自发地发生脱氨基反应。脱氨基反应是氨基被羟基替换，将天冬酰胺和谷氨酰胺分别转化为天冬氨酸和谷氨酸，并且增加了一个负电荷。天冬酰胺和天冬氨酸残基可以发生环化而形成琥珀酰胺。琥珀酰胺的环状结构可以通过两种方式打开：在大多数情况下形成异天冬氨酸，形成的两个亚甲基组成主链；另一种情况是变回天冬氨酸（图 E.3）。脱氨基化反应以及异天冬氨酸的形成都是蛋白质老化的标志。阿尔兹海默症患者脑中的 Ab 肽段就含有超出正常量的异天冬氨酸残基，但目前尚不清楚这是该病的诱因还是结果。真核细胞通常具有一个负责修复的酶——蛋白质 -L- 异天冬氨酸转甲基酶（protein-L-isoaspartyl methyltransferase，PIMT），催化异天冬氨酸转化为正常的天冬氨酸。该酶缺陷则导致蛋白质中异天冬氨酸的积累。

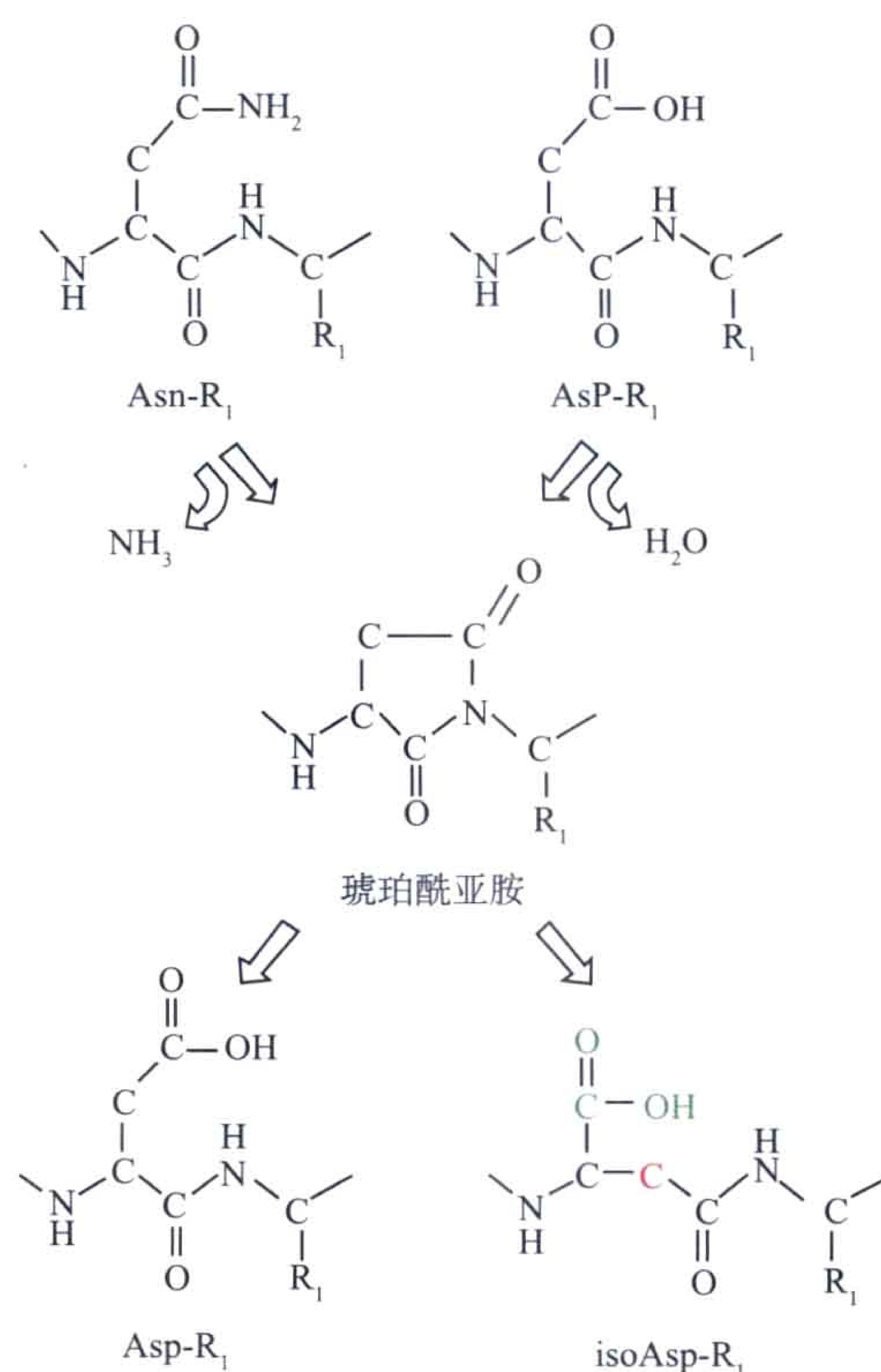


图 E.3　天冬酰胺残基的脱氨基化作用以及天冬氨酸和天冬酰胺环化途径。

若原始氨基酸残基为天冬酰胺，则会发生脱氨基化并增加一个正电荷。到目前为止，PDB 中仅有很少的异天冬氨酸的报道。也有可能这种修饰在很多情况下被忽视了。谷氨酰胺和谷氨酸也能发生上述的反应过程，但发生率较低。

E.4 磷酸化和去磷酸化

蛋白质磷酸化作用作为一种调控机制的研究可以追溯到 Krebs 和 Fischer 对糖原磷酸化酶的研究。磷酸化对细胞内和信号跨膜转导是首要的，很大一部分蛋白质都受到磷酸化的调控。不仅磷酸化非常重要，去除磷酸化反应也同等重要，而去除磷酸基团的过程是由磷酸化酶催化的。胞内蛋白的磷酸化是细胞内各种生理过程的分子开关（详见第十一章）。蛋白激酶家族含有 700 个以上的蛋白质。因此，基因组中很大一部分蛋白质是蛋白激酶或磷酸化酶。蛋白激酶都具有一个约 250 个氨基酸残基的折叠结构。激酶催化的是将 ATP 的 γ 磷酸转移到其底物上的反应。被磷酸化的氨基酸残基可以是酪氨酸、丝氨酸 / 苏氨酸，但是在极少数情况下也有可能是组氨酸和天冬氨酸。由于激酶在信号转导和各种调节机制中起重要作用，所以激酶是药物设计的靶标。

起初，研究人员在激酶的底物中寻找一致的识别序列，但是随着大量的激酶及其底物被鉴定出来，发现在底物中寻找线性的保守识别序列是不可行的，因为激酶与底物的几个模体（motif）都有相互作用。由于磷酸化具有调节功能，这就要求其具有很高的特异性。激酶通过很多方式调节活性。很多激酶自身受到磷酸化激活作用，在这种情况下，这些激酶具有一个活化片段。激酶在开和关两种状态下转换（详见第十一章）。另一种情况是，激酶的若干个结构域需要与激酶的催化结构域相互作用，或者需要多亚基蛋白质的参与。cAMP 依赖的、cGMP 依赖的蛋白激酶 C（AGC 激酶家族）就是被严格调控的。图 E.4 显示了激酶与其底物相互作用的几种形式。

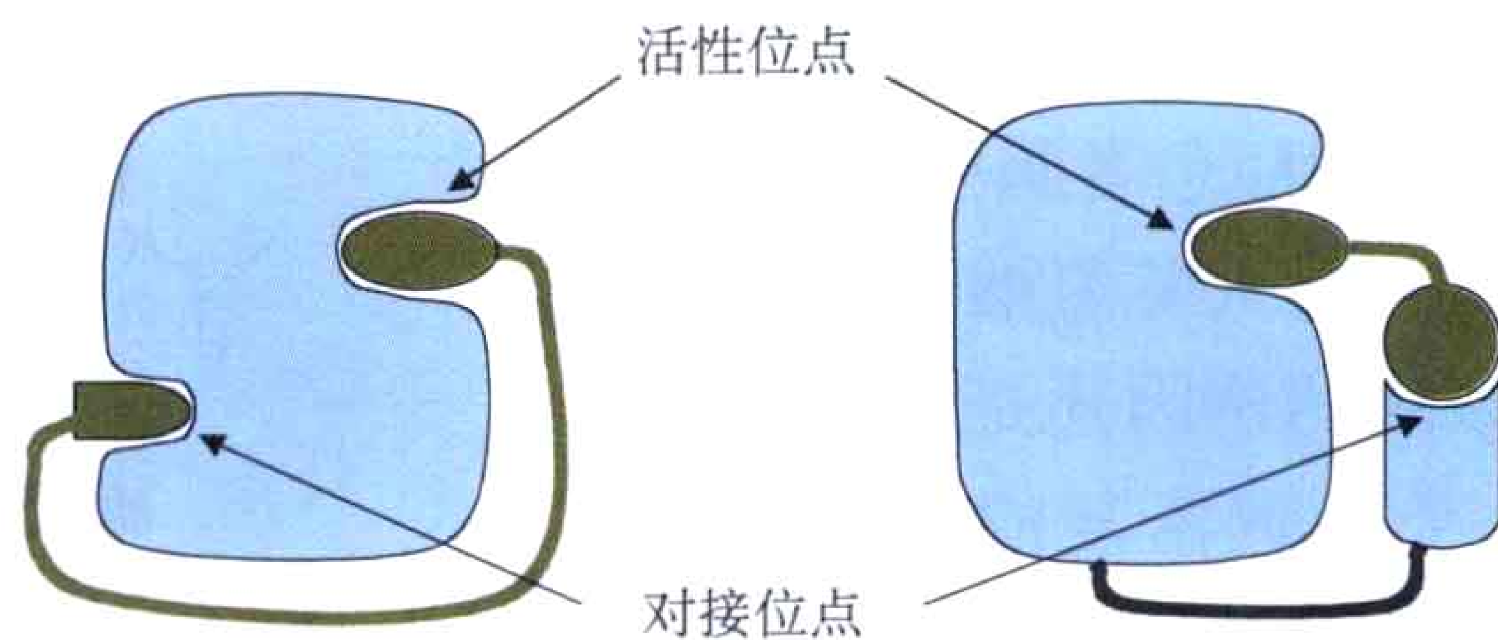


图 E.4 激酶和磷酸酶的调控作用需要通过蛋白间相互作用控制。酶（浅蓝色）需要一定程度的活化，但同时需要识别被磷酸化或去磷酸化部分的结构。这是因为该酶与其底物（绿色）的相互作用是通过一个对接位点（docking site）介导的，这个位点可以是激酶结构域的一部分，或是属于该酶的其他结构域。

蛋白磷酸酶分为几个不同的类型：蛋白丝氨酸 - 苏氨酸磷酸酶、蛋白酪氨酸磷酸酶和蛋白双特异性磷酸酶。它们可能具有一个或几个结构域，而这些结构域可能属于不同的蛋白质家族，其中一部分磷酸酶锚定在膜上或跨膜。所有的蛋白酪氨酸磷酸酶（PTP）都具有一个独特的标志性的模体 $HC(X)_5R$，其催化结构域有大约 280 个氨基酸残基，且其中很多氨基酸残基是不变或高度保守的。

E.5 甲基化作用

甲基化是一种普遍存在的蛋白质、核酸甚至小分子的修饰方式。DNA 中胞嘧啶碱基的甲基化也是基因沉默的一种重要性质。甲基化反应由甲基转移酶介导。*S*- 腺苷甲硫氨酸（S-adenosyl methionine，SAM）是甲基传递载体。精氨酸和赖氨酸可以被甲基化（图 E.5）。赖氨酸的甲基化发生在 ε 氨基上，可以是单甲基化、二甲基化或三甲基化。发生甲基化后赖氨酸的电荷不发生变化。精氨酸也可以发生单甲基化或二甲基化，而二甲基化又存在对称和不对称两种方式。不同的甲基化酶介导不同程度的甲基化。

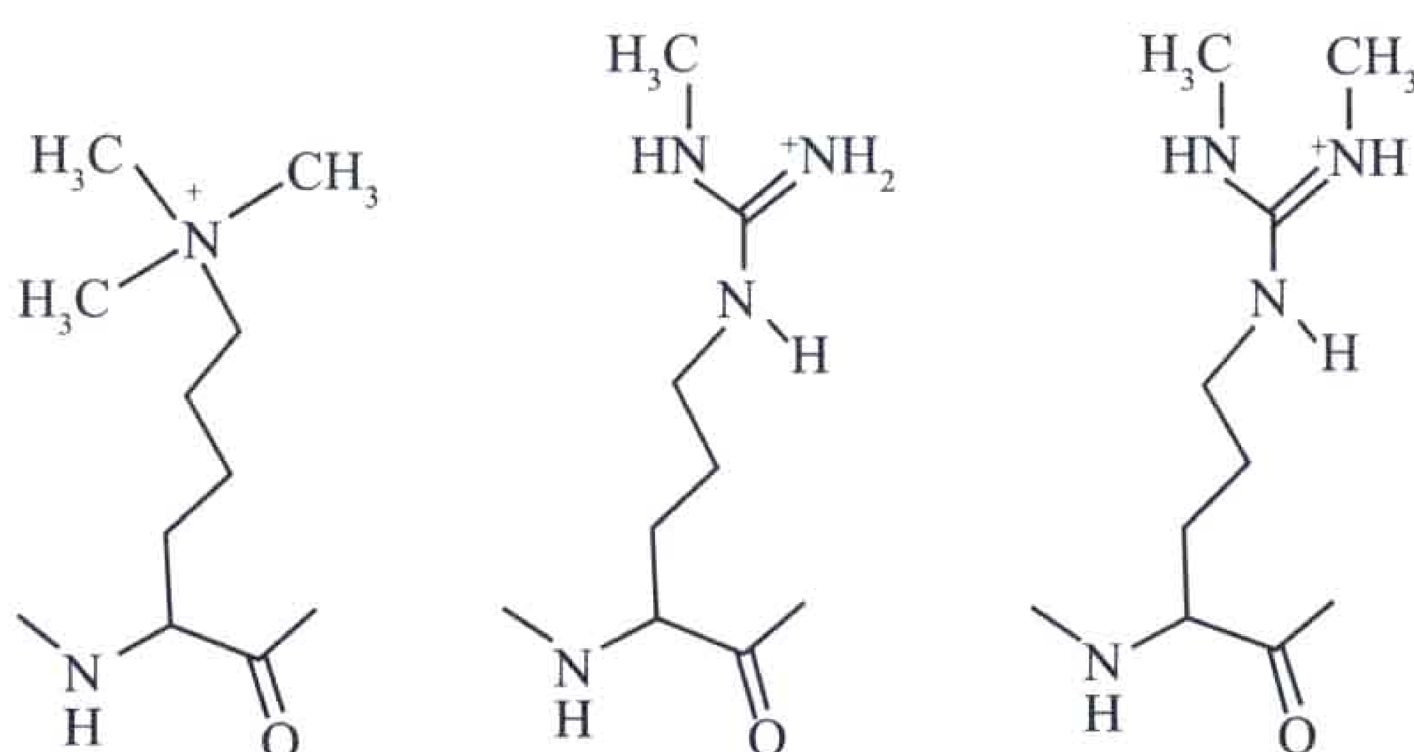


图 E.5 赖氨酸和精氨酸甲基化在蛋白质与核酸相互作用中起重要作用。左边是三甲基赖氨酸，中间和右边分别是单甲基精氨酸和对称的二甲基精氨酸。

甲基化在细菌趋药性和转录调节两个系统中起着重要作用。在细菌趋药性系统中，跨膜受体的甲基化和去甲基化受到甲基转移酶和甲基酯酶的调控。在大肠杆菌中，甲基转移酶为 CheR，甲基酯酶是 CheB。

组蛋白尾部的甲基化及其在转录调控中的作用曾在第七章里单独介绍过。组蛋白的甲基化是由高度特异的组蛋白甲基转移酶介导的。特定组蛋白中的特异性赖氨酸发生甲基化，从而使核小体的排列变得松散。最近，组氨酸甲基酯酶才被鉴定出来，它们也具有与组氨酸甲基转移酶一样的高特异性。

精氨酸的甲基化正受到越来越多的关注。很多蛋白质都会被修饰，包括参与转录的组蛋白和非组蛋白。很多与核酸相互作用的蛋白质都会发生精氨酸甲基化。富含甘氨酸 - 精氨酸的模体常常是这种修饰作用的靶标。

E.6 乙酰化

包括组蛋白在内的很多蛋白质都会被乙酰化修饰。通常，氨基是乙酰化的靶标，被修饰的可以是任何氨基酸的 N 端 α 氨基或赖氨酸的 ε 氨基（图 E.6）。另一种情况

是由组蛋白乙酰转移酶（histone acetyltransferases，HAT）介导的修饰反应，组蛋白乙酰转移酶可以被分为很多家族。TFIID 的亚基和转录机器的中介物（mediator）都是 HAT（详见第七章）。很多 HAT 会与其他蛋白质形成多蛋白复合体而发挥功能。部分 HAT 含有肌动蛋白或肌动蛋白相关蛋白。组蛋白乙酰化的过程是可逆的，因此也存在组蛋白脱乙酰酶（histone deacetylases，HDAC）。组蛋白脱乙酰酶也分为几个家族。

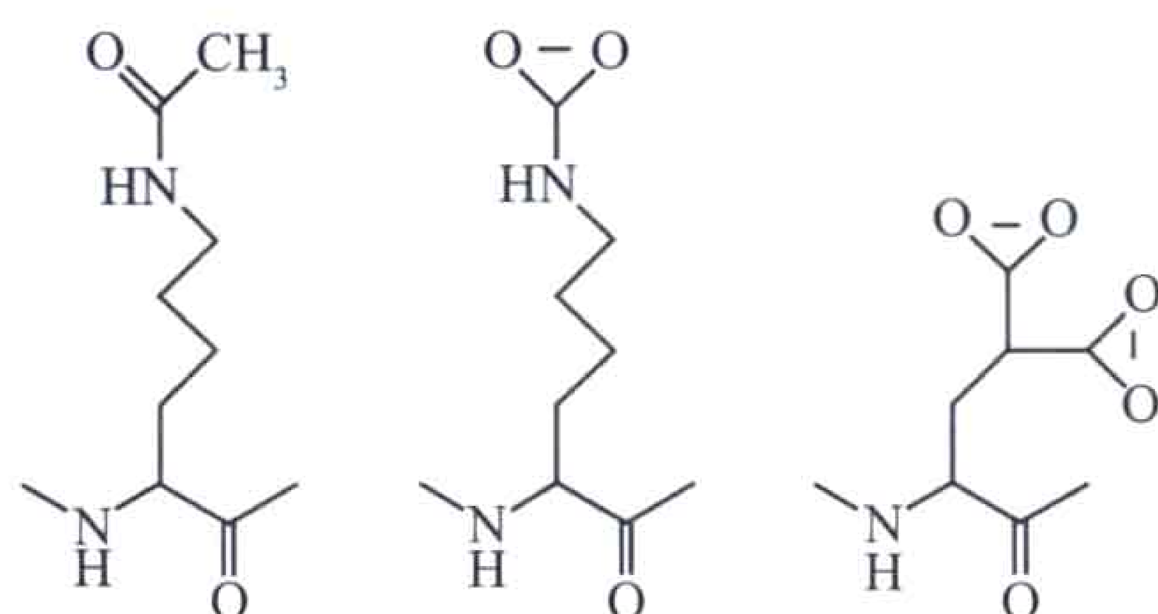


图 E.6　左：乙酰化赖氨酸，这种修饰对组蛋白的修饰非常重要。中：氨基羧基化赖氨酸，在很多活性位点含有两个金属的酶中都发现了这种修饰，这种延长的氨基酸残基可以形成一个“桥”将两个金属原子相连。右：γ- 羧基化谷氨酸（γ-carboxyglutamic acid，Gla），这种修饰广泛的存在于凝血系统中。

E.7　羧基化作用

维生素 K 依赖的蛋白质具有一种特有的修饰残基——γ- 羧基化谷氨酸（Gla，γ-carboxyglutamic acid）。这种特殊的氨基酸残基是由 γ 谷氨酰基羧化酶催化的谷氨酸翻译后羧基化形成的。被修饰的蛋白质中需要有与该酶紧密结合的前肽。除了还原性的维生素 K 之外，CO_2 和 O_2 也是必需的。凝血系统中的很多蛋白质以及一些参与成骨过程的酶都是维生素 K 依赖的。凝血蛋白 N 端含有 Gla 的结构域，长约 45 个氨基酸残基，这种结构域含有 9~12 个 Gla 残基。这种羧基化修饰使得蛋白质可以结合 Ca^{2+}。钙是凝血系统和成骨过程的关键元素。在一些情况下，Gla 结构域结合 7 个钙离子。由谷氨酸变成 γ - 羧基化谷氨酸并结合钙的过程可引发一个很大的构象变化。带两个电荷的 Gla 残基指向钙结合结构域的内部，导致了疏水氨基酸残基的外露和蛋白质与膜的结合。

在二磷酸核酮糖氧合酶 / 羧化酶、核酮糖 -1,5- 二磷酸羧化 / 加氧酶（rubisco）、内酰胺酶以及尿素酶中也发现了赖氨酸羧基化或氨基羧基化的现象。在二氢乳清酸酶（dihydroorotase）和二氢嘧啶酶（dihydropyrimidinase）中（详见第十六章），一个金属中心的两个锌离子通过一个羧基化赖氨酸残基相连。这种修饰可能是由赖氨酸与金属中心与 CO_2/HCO_3^- 反应形成的。

E.8　十四烷酰化、异戊二烯化及膜锚定

疏水性基团对蛋白质的特异性修饰可以使蛋白质锚定在膜上。某些膜结合是通过十四烷酰化、异戊二烯化或糖基磷脂酰肌醇（glycosylphosphatidylinositol，GPI）锚定形成的。

十四烷酰化是一种存在于真核生物或病毒中对甘氨酸 α 氨基的不可逆翻译后酶促修饰反应。这种修饰会导致蛋白质与蛋白质在膜上的相互作用，起到调节和信号转导的作用。

蛋白质的异戊二烯化是对靶标蛋白 C 端的半胱氨酸进行法尼基（farnesyl）或香叶草 - 香叶草基（geranyl-geranyl）的酶促转移。被修饰的酶的 C 端具有一个 CaaX 保守序列，其中“a”表示脂肪族氨基酸残基，“X”决定是哪种酶参与催化反应。香叶草基有 10 个碳的骨架，法尼基 15 个碳，而香叶草 - 香叶草基有 20 个碳。G 蛋白家族，如 Ras 和 Rab（详见第五章和第十一章）会发生法尼基化。异戊二烯化的抑制剂由于具有治疗癌症和寄生虫病的潜力而受到广泛关注。

GPI 锚定会发生在蛋白的 C 端。这两个脂肪酸部分将蛋白质锚定在细胞膜上。蛋白质的这个基团具有一个疏水的 C 端，蛋白质通过这个基团与内质网结合。在内质网上，这个疏水的末端被 GPI- 锚定代替，而且蛋白质通过囊泡被转运至高尔基体。最后，蛋白质附着在细胞膜的外层。锚定可以被磷脂酶切除从而释放蛋白质。

E.9　糖基化

糖基化是指在蛋白质上共价结合一个或多个糖基。糖基化可以是酶促反应，这种糖基化发生在特异的位点；也可以是非酶促反应（糖化作用），而这种情况下会损坏酶的功能。高达一半以上的蛋白质都会受到糖基化修饰，这使得糖基化成为最普遍的翻译后修饰作用。首先，糖基化发生在真核细胞和古细菌中，几乎所有的分泌蛋白或真核生物膜结合蛋白都是糖基化的，而大部分胞质蛋白和核蛋白都不是糖基化的。糖基化是一个很大的领域，在此仅仅提供一个简短的总结。

蛋白质存在 *O*- 或 *N*- 两种糖基化形式。*O*- 糖基化发生在丝氨酸和苏氨酸残基的侧链上。*N*- 糖基化发生在保守序列 NXS/T 的天冬酰胺残基上，其中 X 不能是脯氨酸。丝氨酸或苏氨酸的羟基增加天冬酰胺酰胺基的亲核性。

蛋白质的糖基化过程非常复杂，可以在蛋白质上添加不同种类的糖基（图 E.7），而且糖基化可以是单个单糖修饰，也可以是多个单糖直链或支链相连的修饰。一个蛋白质中可能有一个或几个氨基酸残基会发生糖基化。有一系列的酶参与了蛋白质的糖基化过程，它们被称为糖基转移酶。也存在一系列的糖基水解酶。这两种酶占去了已

知基因组编码蛋白的 1% ~3%。目前，与糖基化相关的酶被分为 200 多个家族（http://afmb.cnrs-mrs.fr/CAZY/）。糖基化修饰暴露在蛋白质表面，而且倾向于处于高度柔性的状态。对于晶体学工作，通常倾向于去除蛋白质表面的寡糖链，而至今还没有观察到去除寡糖链的蛋白质发生构象变化。

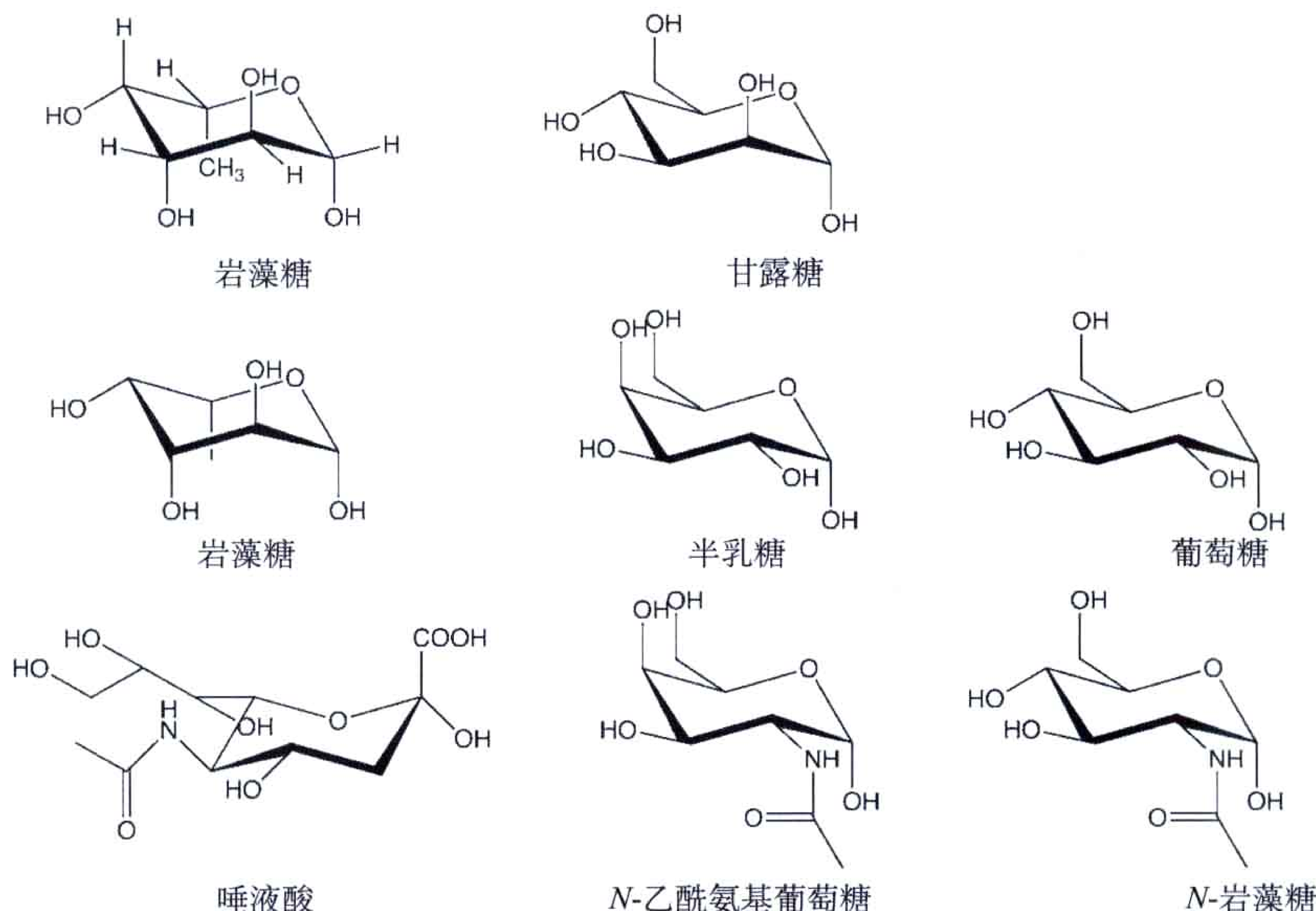


图 E.7　一些可以附加在蛋白质上的糖的类型。左上是岩藻糖，其分子式中的所有的氢原子都显示出来。下，同样是岩藻糖，但是直接与糖环相连的氢没有显示出来。

O- 连接寡糖一般比较短（1~4 个单糖残基），它们与丝氨酸或苏氨酸的羟基通过 *N*-乙酰半乳糖胺相连。这种寡糖链中的单糖残基是通过不同的糖基转移酶一个一个加到蛋白质上的。催化这些反应的糖基转移酶通常是膜整合蛋白，它们对单糖核苷酸分子和受体（蛋白质或是预先结合的糖）都具有选择性，而且它们对于新的糖基加在哪个羟基上也是特异性的。

在 *N*- 连接的寡糖中，*N*- 乙酰氨基葡萄糖结合在天冬酰胺侧链氮原子上（图 E.8）

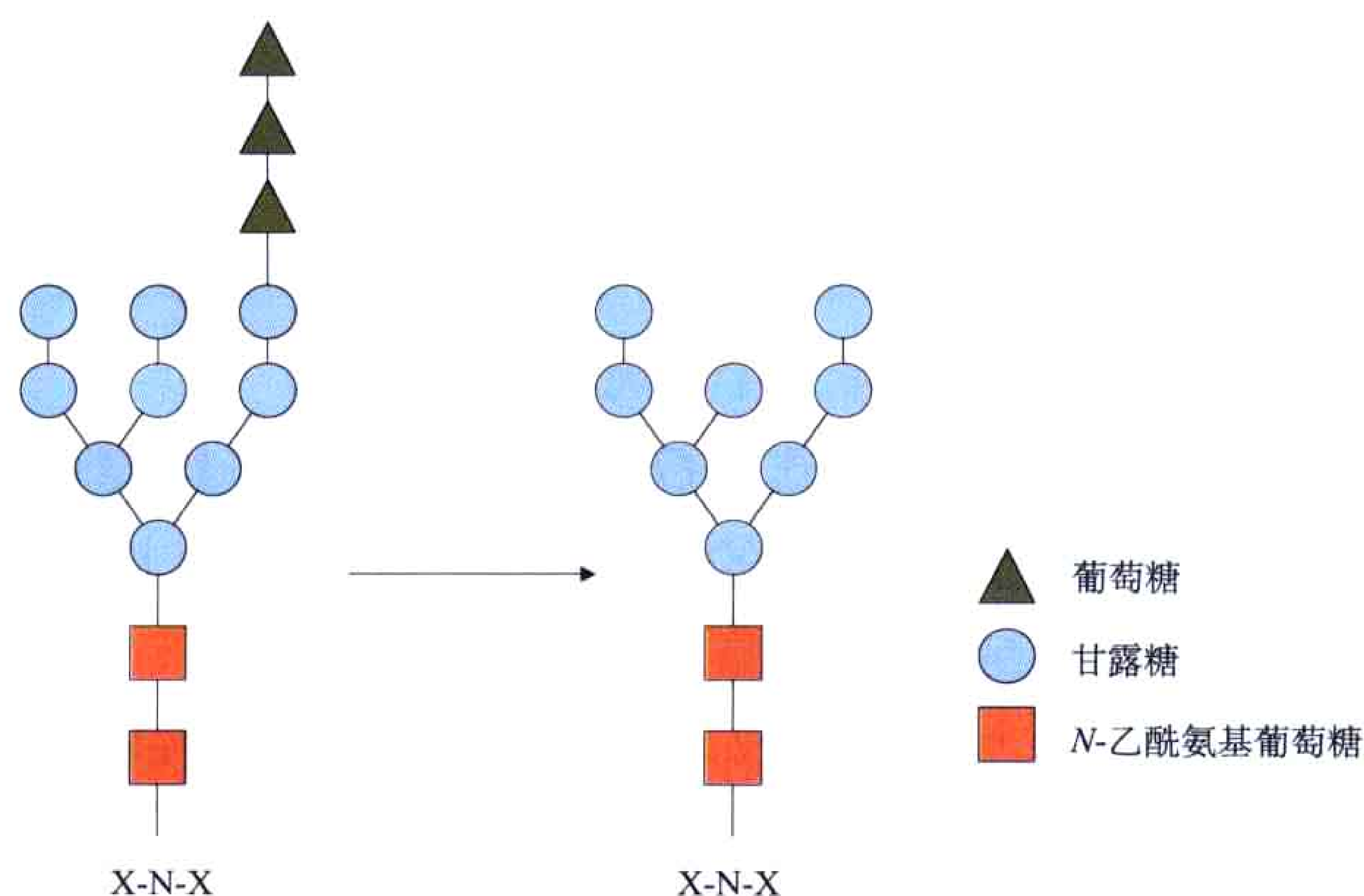


图 E.8　天冬酰胺侧链 *N*- 连接寡糖修饰的例子，其中含有三种不同的单糖共 14 个残基。在发生进一步的特异性修饰之前，这个寡糖链通过多步反应被削减至 10 个单糖残基。

这种修饰基团是一个大的分支的寡糖链，它原本结合在固定在膜上的长萜醇分子上，其组成是 (葡萄糖 $)_3$(甘露糖 $)_9$(*N*- 乙酰葡萄糖 $)_2$。被转到蛋白质上之后，三个葡萄糖和一个特异性的甘露糖被切除。这种修饰发生在内质网上，而进一步的修饰反应则发生在高尔基体上。

蛋白质糖基化修饰功能是多方面的：在一些情况下，它们帮助蛋白质正确折叠或是活性中心的一部分；另一些情况下，它们稳定蛋白质使蛋白质不被降解；有时寡糖链也会参与细胞与细胞的联系。

E.10　交联作用

在很多情况下，蛋白质内和蛋白质间的氨基酸残基存在着交联现象。交联反应最常见的类型是形成二硫键。芳香族也经常参与形成交联，它们可以和芳香族基团或其他氨基酸残基形成交联。例如，二酪氨酸的形成，它是个荧光分子（图 E.9）。这种交联可能是通过酪氨酸或苯丙氨酸的氧化形成的，它在很多结构蛋白中天然存在，可以增加蛋白质的结构稳定性和弹性，而且二酪氨酸不受蛋白水解酶作用。

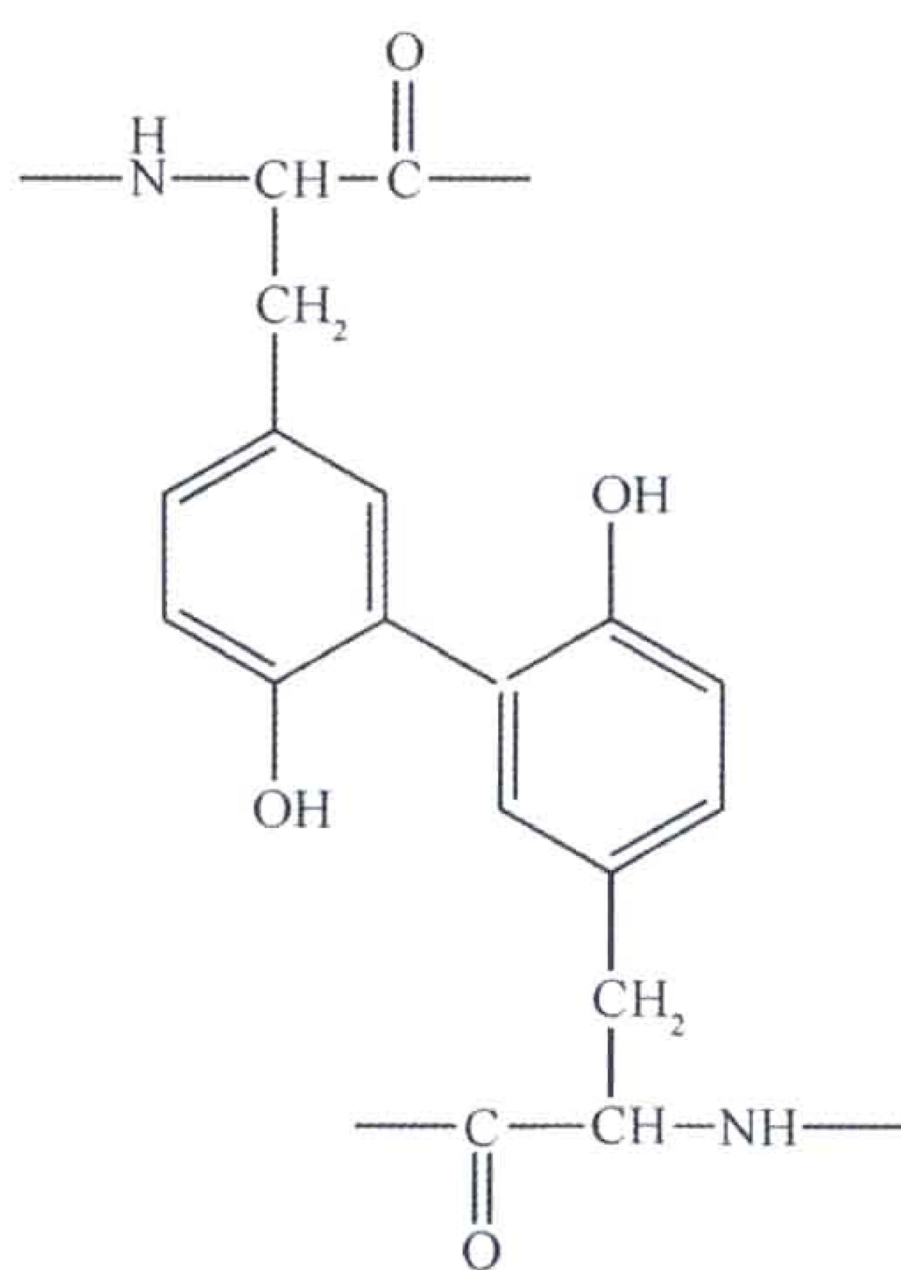


图 E.9　两个酪氨酸残基之间形成交联的一种形式。还有其他一些形式也被发现。

异肽键是另一种形式的蛋白质内或蛋白质间的交联反应。一般来说，异肽键的形成是由转谷氨酰胺酶催化形成，发生在谷氨酰胺残基的甲酰胺基团与另一个氨基酸的顶端氨基，一般是赖氨酸残基的 ε - 氨基之间。转谷氨酰胺酶通常是 Ca^{2+} 依赖的。这种高分子质量的产物对蛋白酶水解是高抗的。用转谷氨酰胺酶处理食品蛋白会增加成胶性和吸水性。来自噬菌体 HK97 衣壳蛋白相邻亚基的一个天冬酰胺和一个赖氨酸残基可以通过自催化形成另一种异肽键。交联作用通过使 5 个或 6 个亚基形成环状结构，

使各亚基发生联系。最终，衣壳蛋白形成类似“奥运五环”的结构并形成一个巨大分子质量的复合体。这种交联在衣壳蛋白大规模运动时被引发，稳定噬菌体高度扩张的形式，在这种扩张的形式中噬菌体的衣壳蛋白是很薄的。

另外，也存在一些非正常的交联类型。在海洋生物中发现的绿色荧光蛋白（GFP）中存在的发色团交联已经发展成为在分子和细胞生物学中极有价值的工具。GFP 由一个中心包裹了一个 α 螺旋的 11-stranded 的 β 桶组成，其发色团是由三个保守氨基酸残基 Ser65、Tyr66 和 Gly67 的主链自发环化形成的（图 E.10）。类似的主链自发环化作用在其他一些例子中也曾发现。主链交联而形成的咪唑啉酮环与酪氨酸的苯酚基通过亚甲基桥相连，然后该苯酚基和咪唑啉酮环发生完全的 π 共轭。突变发色团中或附近的氨基酸残基可以改变荧光蛋白的颜色，从而发展出 GFP 家族的其他蛋白质。将丝氨酸替换为苏氨酸可以增加荧光亮度。

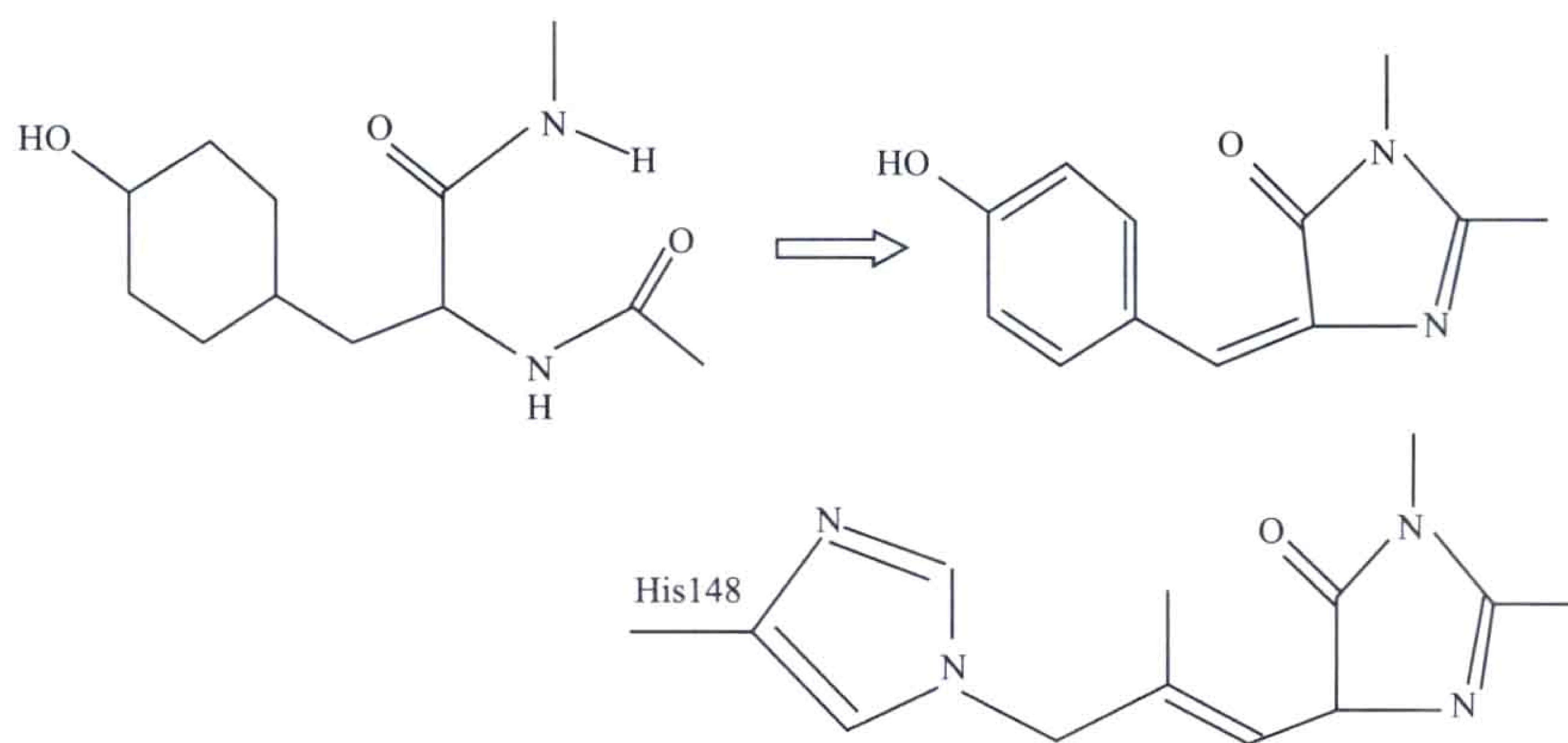


图 E.10 上图：GFP 的 Tyr66 发色基团的形成。下图：酪氨酸突变为亮氨酸的突变体中，与相邻 His148 的反应形成黄色荧光蛋白。

其后还发现了另一些蛋白质中的发色或荧光基团是由于翻译后修饰自催化作用而产生的。

E.11 活性位点的修饰作用和交联反应

在很多例子中，修饰反应位于酶的活性位点。在已被发现的全活性位点范围被修饰的酶中，一些修饰作用是交联反应。这种侧链的共价交联形成了新的具有催化功能的化学结构。

细胞色素氧化酶的晶体学研究中发现，一个与血红素 α_3-Cu_B 中心的 CuB 中的铜原子配位的组氨酸残基距离相邻的酪氨酸非常近，所以认为该组氨酸与该酪氨酸形成共价交联，这已经为随后的化学实验所证实（图 E.11）。这种自发的交联在相关的酶中

是保守存在的。这种修饰作用导致了酪氨酸残基的 pK_a 发生改变，同时也改变了该酶的氧化还原电位。π 共轭系统允许苯酚基和咪唑环之间发生很大的电子重新定位。

在半乳糖氧化酶中发现了一种酪氨酸与半胱氨酸之间形成的交联作用，而且铜离子也参与其中，但是酪氨酸的羟基是铜的一个配体。

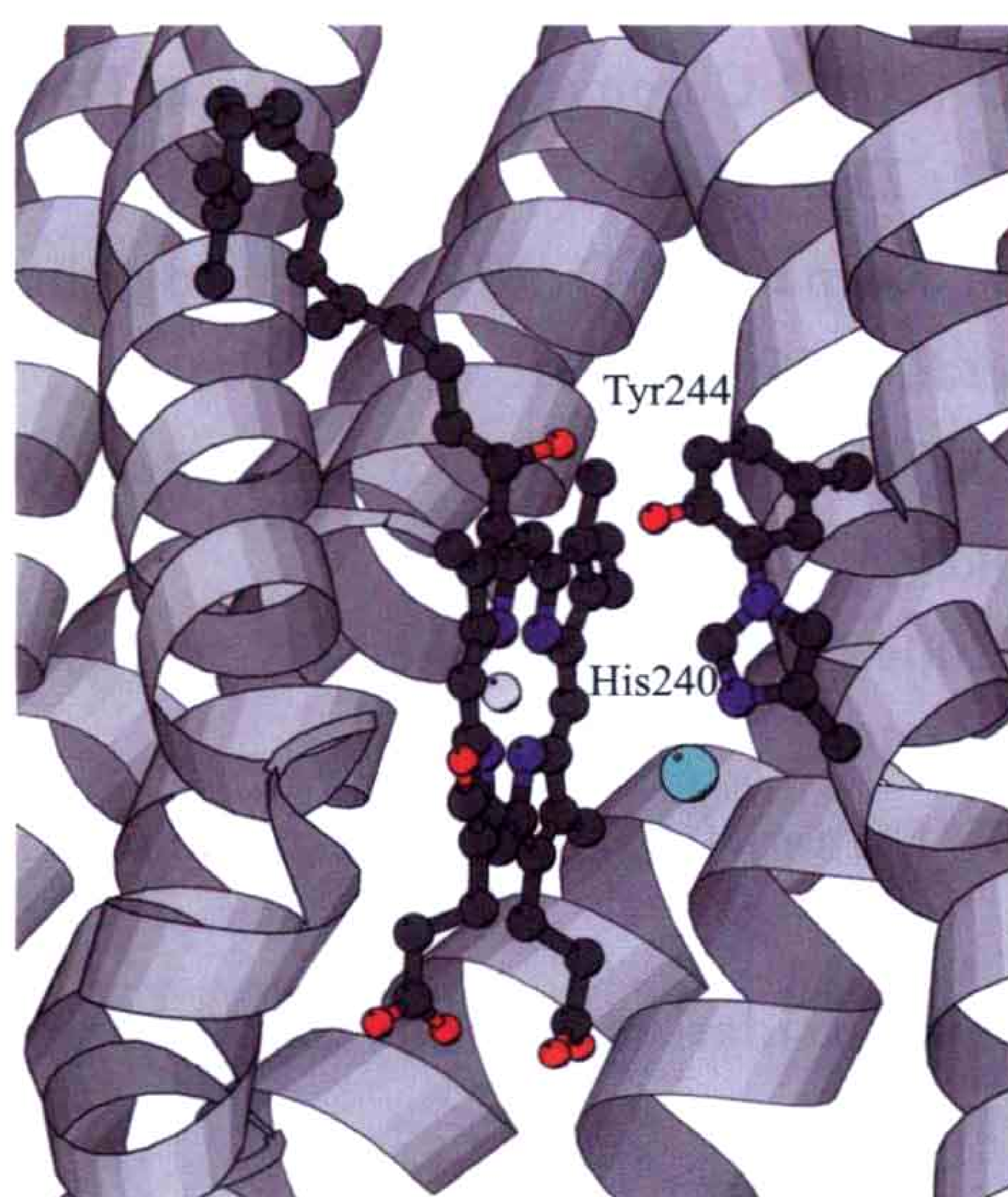


图 E.11　位于细胞色素 c 氧化酶的血红素 α_3-Cu_B 中心的酪氨酸与组氨酸之间的共价连接作用。铜原子显示为蓝绿色（PDB：2EIK）。

延伸阅读

原始文献

Lohkamp B, Andersen B, Piskur J, Dobritzsch D. (2006) The crystal structures of dihydropyrimidinases reaffirm the close relationship between cyclic amidohydrolases and explain their substrate specificity. *J Biol Chem* **281**: 13762-13776.

Yoshikawa S, Shinzawa-Itoh S, Nakashima R, Yaono R, Yamashita E, Inoue N, Yao E, Fej MJ, Libeu CP, Mizushima T, Yamaguchi H, Tomizaki T, Tsukihara T. (1998) Redox-coupled crystal structural changes in bovine heart cytochrome oxidase. *Science* **280**: 1723-1729.

Taylor TC, Andersson I. (1997) The structure of the complex between Rubisco and its natural substrate ribulose 1,5-bisphosphate. *J Mol Biol* **265**: 432-444.

Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**: 2129-2133.

综述文章

Dalle-Donne I, Scaloni A, Glustarini D, Cavarra E, Tell G, Lungarella G, Colombo R, Rossi R, Milzani A. (2005) Proteins as biomarkers of oxidative/nitrosative stress in diseases: The contribution of redox proteomics. *Mass Spect Rev* **24**: 55-99.

Ehrenhofer-Murray AE. (2004) Chromatin dynamics at DNA replication, transcription and repair. *Eur J Biochem* **271**: 2335-2349.

Gold MG, Barford D, Komander D. (2006) Lining the pockets of kinases and phosphatases. *Curr Op Struct Biol* **16**: 693-701.

Hanson K, Stenflo J. (2005) Post-translational modifications in proteins involved in blood coagulation. *J Thromb Haemost* **3**: 2633-2648.

Helenius A, Aebi M. (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **73**: 1019-1049.

Hirsch AKH, Fischer FR, Diederich F. (2007) Phosphate recognition in structural biology. *Angew Chemie* **46**: 338-352.

Nystrom. (2005) Role of oxidative carbonylation in protein quality control and senescence. *EMBO J* **24**: 1311-1317.

Oakley NM, van der Donk WA. (2000) Novel cofactors via post-translational modifications of enzyme active sites. *Chem Biol* **7**: R159-R171.

Peter-Katalinic′ J. (2005) O-glycosylation of proteins. *Meth Enzymol* **405**: 139-171.

Reményi A, Good MC, Lim WA. (2006) Docking interactions in protein kinase and phosphatase networks. *Curr Op Struct Biol* **16**: 676-685.

Shimizu T, Matsuoka Y, Shirazawa T. (2005) Biological significance of isoaspartate and its repair system. *Biol Pharm Bull* **28**: 1590-1596.

Tiganis T, Bennett AM. (2007) Protein tyrosine phosphatase function: The substrate perspective. *Biochem J* **402**: 1-15.

Tsien RY. (1998) The green fluorescent protein. *Annu Rev Biochem* **67**: 509-544.

附录 F

与结构生物学相关的诺贝尔奖得主

结构生物学的历史可以从诺贝尔奖反映出来，自从开奖伊始，诺贝尔物理学奖（P）、化学奖（C）以及生理与医学奖（M）的获得者们就以各种方式贡献于结构生物学。我们按照年代顺序来简短描述对结构生物学做出卓越贡献的获得诺贝尔奖的科学家们，这里仅限于最关键的贡献，关于这些获奖者的详细信息及图片资料见诺贝尔网站：www.nobel.org 。

1901 年 伦琴 Wilhelm Conral Rontgen (P)：由于发现 X 射线而获得第一个诺贝尔物理学奖。在许多有用的特性中，X 射线被不同物质的散射性质得到广泛的研究应用。这特别导致了 X 射线晶体衍射的发现，第一个蛋白质晶体的 X 射线衍射实验由霍奇金（Dorothy Crowfoot Hodgkin）在 1934 年完成，所用蛋白质是胃蛋白酶。

1914 年 劳埃 Max von Laue (P)：发现晶体的 X 射线衍射。此工作由理论物理学家劳埃与两个年轻科研工作者李平（Paul Knipping） 和弗里德里希（ Walter Friedrich）共同完成。

1915 年 亨利・布拉格爵士 Sir William Henry Bragg，劳伦斯・布拉格爵士 Sir William Laurence Bragg (P)：由于晶体结构的 X 射线分析及解析。食盐（NaCl）的结构是他们得到的第一个晶体结构。

1926 年 斯维德伯格 Theodor (The) Svedberg (C)：由于分散体系的研究荣获诺贝尔化学奖，特别是他发展了超速离心方法分析蛋白质大分子，如血红蛋白的分子质量。

1946 年 萨姆纳 James Batcheller Sumner，诺思罗普 John Howard Northrop 和斯坦利 Wendell Meredith Stanley (C)：由于发现结晶蛋白酶，制备活性状态的酶和病毒蛋白质。萨姆纳确立了酶是蛋白质，特别是他发现并结晶了一些酶，例如他纯化及结晶了脲酶，这也确立了酶和蛋白质具有确定的三维结构。诺思罗普随后纯化并结晶了胃蛋白酶、胰蛋白酶原、胰蛋白酶、Kunitz 胰蛋白酶抑

制剂、糜蛋白酶原、糜蛋白酶、羧肽酶、核酸酶及己糖激酶。斯坦利主要致力于病毒工作，分离、纯化并结晶了烟草花叶病毒和番茄丛矮病毒。

1954 年 鲍林 Linus Carl Pauling (C)： 由于研究化学键的本质及蛋白质结构。 他发展了蛋白结构模型构建方法，提出 α 螺旋和 β 片层结构是蛋白质中两种最稳定的构象。

1958 年 桑格 Frederick Sanger (C)：发明蛋白质的氨基酸序列测序方法，特别是胰岛素的测序工作。桑格后来还发展出了 DNA 序列测定方法，并因此于 1980 年再次获得诺贝尔化学奖。

1962 年 帕鲁兹 Max Ferdinand Perutz 和 肯德鲁 John Cowdery Kendrew (C)： 由于发展蛋白质晶体学方法及首次测定了血红蛋白与肌红蛋白的晶体结构。肯德鲁的肌红蛋白测到 2 Å 分辨率，帕鲁兹的血红蛋白结构则在较低的分辨率。帕鲁兹发展了多重原子同晶置换方法确定衍射数据的相位。电子密度图显示了大多数原子的确定位置及大部分的 α 螺旋结构，从血红素基团及铁原子的位置，可以导出氧原子结合的机理。从这些最先测定的蛋白质结构中可以得到的一个很重要的结果是血红蛋白与肌红蛋白的结构很相似，从而建立了共同蛋白折叠的概念。

1962 年 克里克 Francis Harry Compton Crick，沃森 James Watson，威尔金斯 Maurice and Maurice Hugh Frederick Wilkins (M)：由于确定了核酸（DNA）的分子结构及其在生物信息传递中的重要性。威尔金斯和富兰克林（Rosalind Franklin）实验得到纤维衍射数据，克里克和沃森得到能够解释数据的原子模型。

1964 年 霍奇金 Dorothy Crowfoot Hodgkin (C)：由于测定青霉素、维生素 B12 等重要生物分子的晶体结构测定。 她后来还确定了胰岛素的结构。

1972 年 安芬森 Christian Borhmer Anfinsen，摩雷 Stanford Moore，以及 斯坦 William H．Stein(C)：发现蛋白质折叠的基本原则及研究酶化学的基本理论。 安芬森研究了酶的折叠，特别是核酸酶及金葡菌核酶。他指出一个蛋白质的折叠仅仅依赖于氨基酸序列。摩雷及斯坦获奖是由于发展了研究及表征蛋白质与酶，特别是 RNA 核酶及 DNA 核酶的化学方法。

1978 年 米切尔 Peter D. Mitchell (C)：由于研究解释生物系统中的能量转移过程的化学渗透理论。

1980 年　伯格 Paul Berg；吉尔伯特 Walter Gilbert 和桑格 Frederick Sanger(C)：建立 DNA 一级结构的化学和生物分析法。

1982 年　克卢格 Aaron Klug (C)：发展电子晶体学（电子显微镜）方法测定具有重要功能的核酸蛋白复合物的三维结构。此方法最早用于病毒的结构研究，后来发展到生物化学及分子生物学中其他的基本生物大分子结构问题。克卢格还曾领导最早确定 tRNA 晶体结构的一个研究组。

1985 年　霍普特曼 Herbert A. Hauptman 和卡尔勒 Jerome Karle (C)：发展直接法解析晶体结构。该方法主要用于小分子晶体结构研究，也部分用于大分子晶体结构解析。

1988 年　戴森霍弗 Johann Deisenhofer，休伯 Robert Huber，以及米歇尔 Hartmut Michel (C)：光合作用反应中心（第一个膜蛋白）的三维晶体结构研究。这是膜蛋白结构及光合作用机理研究方面的一次巨大飞跃。

1997 年　博耶 Paul D. Boyer 和沃克 John E Walker (C)：阐明“能量分子”三磷酸腺苷（ATP）的形成过程（ATP 合酶）；斯科 Jens Skou (C)：在细胞中首次发现跨膜离子转运钠 - 钾 ATP 酶 (Na^+/K^+-ATPase)。

2002 年　伍斯里奇 Kurt Wüthrich (C，三人之一)：发展核磁共振光谱学技术测定生物大分子的三维结构。

2003 年　阿格雷 Peter Agre (C)：发现水通道；麦金农 Roderick MacKinnon (C)：离子通道的结构和机理研究。

2006 年　科恩伯格 Roger D. Kornberg (C)：由于真核细胞转录的分子机理研究。酵母的 RNA 聚合酶及其相互作用的众多转录因子、DNA 模板和核苷酸复合物结构在不同状态下得以阐明。

译者注：

2009 年，拉马克里希南 Venkatraman Ramakrishnan, 尤纳斯 Ada Yonath, 施泰茨 Thomas Steitz (C)：由于细菌核糖体结构和功能的研究。

2012 年，莱夫科维茨 Robert J. Lefkowitz，克比尔卡 Brian K. Kobilka (C)：由于 G 蛋白偶联受体（GPCR）的结构与功能研究。

索引